La Inteligencia Artificial en el proceso penal de instrucción español: posibles beneficios y potenciales riesgos

Carlota Cuatrecasas Monforte

TESIS DOCTORAL

Título	La Inteligencia Artificial en el proceso penal de instrucción español: posibles utilidades y potenciales riesgos.
Realizada por	Carlota Cuatrecasas Monforte
en el Centro	Facultad de Derecho Esade
y en el Departamento	Derecho
Dirigida por	Dr. David Velázquez Vioque Dr. Eloy Velasco Núñez

ABREVIATURAS

ACLU	The American Civil Liberties Union
ACM	Association for Computer Machinery
ACNUR	Oficina del Alto Comisionado de Naciones Unidas para los Refugiados
AEPD	Agencia Española de Protección de Datos
AP	Audiencia Provincial
Art. (Arts.)	Artículo (s)
AVPD	Agencia Vasca de Protección de Datos
BAAI	Beijing Academy of Artificial Intelligence
BOE	Boletín Oficial del Estado
CE	Comisión Europea
CGPJ	Consejo General del Poder Judicial
CP	Código Penal
EEUU	Estados Unidos
Ej.	Ejemplo
IA	Inteligencia Artificial

INE	Instituto Nacional de Estadística
LECrim.	Ley de Enjuiciamiento Criminal
LO	Ley Orgánica
LOPJ	Ley Orgánica del Poder Judicial
OCDE	Organización para la Cooperación y el Desarrollo Económico
OECD	Organization for Economic Co-operation and Development
ONU	Organización de las Naciones Unidas
Op. Cit.	Obra Citada
OTAN	Organización del Tratado Atlántico Norte
P.	Page
Pág./Págs.	Página (s)
PE	Parlamento Europeo
RD	Real Decreto
RDL	Real Decreto Legislativo
SAP	Sentencia de la Audiencia Provincial
s.f.	Sin fecha

STC	Sentencia del Tribunal Constitucional
STS	Sentencia del Tribunal Supremo
TC	Tribunal Constitucional
TS	Tribunal Supremo
UE	Unión Europea
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
UNICRI	United Nations Interregional Crime and Justice Research Institute
VALCRI	Visual Analytics for sense-making in Criminal Intelligence Analysis
Vol.	Volumen

ÍNDICE

1.- INTRODUCCIÓN

Ya en el siglo XIX Mariano José de Larra (1809-1837), uno de los grandes del romanticismo político español, plasmó con sátira en su conocido artículo titulado "Vuelva usted mañana" [1] la realidad del funcionamiento de los servicios en España y, en concreto de la Administración Pública.

Y es que en tal relato exponía con gran sarcasmo cómo un extranjero llegaba a nuestro país decidido a realizar una inversión, con la idea de dejar todos los trámites terminados en quince días, y acababa desesperado al cabo de los meses por la lentitud y la ineficiencia de todos aquellos que debían prestarle colaboración, sobre todo administrativa.

En tal sentido, considero muy ilustrativo traer a colación uno de los brillantes pasajes del mencionado artículo que, sin duda y, por desgracia, no solo reflejan la realidad de la Administración española del siglo XIX, sino también la de la actual:

"*Vuelto de informe se cayó en la cuenta en la sección de nuestra bendita oficina de que el tal expediente no correspondía a aquel ramo; era preciso rectificar este pequeño error; pasose al ramo, establecimiento y mesa correspondiente, y hétenos caminando después de tres meses a la cola siempre de nuestro expediente, como hurón que busca el conejo, y sin poderlo sacar muerto ni vivo de la huronera. Fue el caso al llegar aquí que el expediente salió del primer establecimiento y nunca llegó al otro.*
-De aquí se remitió con fecha de tantos -decían en uno.
-Aquí no ha llegado nada -decían en otro.
-¡Voto va! -dije yo a Monsieur Sans-délai, ¿sabéis que nuestro expediente se ha quedado en el aire como el alma de Garibay, y que debe de estar ahora posado como una paloma sobre algún tejado de esta activa población? Hubo que hacer otro. ¡Vuelta a los empeños! ¡Vuelta a la prisa! ¡Qué delirio!
-Es indispensable -dijo el oficial con voz campanuda-, que esas cosas vayan por sus trámites regulares. (…)

[1] De Larra, 1833.

-¿Para esto he echado yo mi viaje tan largo? ¿Después de seis meses no habré conseguido sino que me digan en todas partes diariamente: «Vuelva usted mañana», y cuando este dichoso «mañana» llega en fin, nos dicen redondamente que «no»? (…)"

Bien es sabido que, en la actualidad, la Administración de Justicia es la Administración Pública peor valorada por la población española, y es que solo uno de cada cinco españoles aprueba su funcionamiento, según una encuesta publicada por el CIS en el año 2020[2], lo cual, bajo mi punto de vista, resulta inaceptable.

Inaceptable porque está claro que algo falla. Y es que si bien la mayor parte de los profesionales de la Administración de Justicia trabajan a destajo para que los procedimientos salgan adelante de la mejor forma posible, lo cierto es que la justicia no funciona. Es lenta, es ineficiente y, por ende, no ofrece un servicio de calidad.

Y ello, desde luego, no es un problema que solo sufran los ciudadanos que deben esperar años, muchas veces con angustia, que sus asuntos queden resueltos. Ello es un problema que también afecta a las personas que trabajan en la Administración, que en la mayoría de casos se ven absolutamente sobrepasadas por la carga de trabajo tan elevada que tienen, que es humanamente imposible de digerir. Y, evidentemente, ante tal problemática, perdemos todos.

Tal circunstancia, sin duda, es la que, tras mis primeros años de trayectoria profesional, me impulsó a pensar en posibles soluciones. Soluciones que veía indispensables para poder seguir dedicándome a mi profesión, la judicatura, puesto que las actuales condiciones de trabajo no resultan (o no deberían resultar) sostenibles, sobre todo en un mundo en el que tenemos a nuestro alcance infinidad de nuevas tecnologías que pueden resultar de enorme utilidad para mejorar el servicio que da la Administración. Y entiendo que, en la posición de servidora pública que ostento, ello no es una mera iniciativa de "supervivencia" o "buena voluntad", ello es una auténtica responsabilidad.

[2] Véase Castro, 2020.

No obstante, lo cierto es que, por desgracia, no todos los miembros de la Administración de Justicia, por unos motivos o por otros, cuentan con la misma actitud para fomentar y aceptar la introducción de cambios, especialmente tecnológicos. Y, en relación con ello, entiendo oportuno traer a colación lo manifestado por Carlos Delgado Romero, Jefe del Laboratorio de Acústica Forense de la Comisaría General de Policía Científica del Cuerpo Nacional de Policía, que parafraseando al filósofo americano Thomas Khun, dispuso: "*Hago lo que puedo y lo mejor que puedo con los recursos de los que dispongo; a ver si ahora que tengo todo organizado viene alguien y me dice que he de cambiar mi manera de trabajar*."[3]

Y tal actitud, desde luego, es una limitación para la evolución y comporta un verdadero retraso, ya que las personas que la ostentan no solo no aportan ideas novedosas para mejorar el servicio que presta la Administración de Justicia, sino que además muestran su frontal oposición a cualquier novedad que pueda suponer cambios, por lo que entorpecen el camino del progreso.

En cualquier caso, "nunca llueve a gusto de todos" y ello, sin duda, es algo que se tiene que aceptar, puesto que es inherente a la diversidad humana. No obstante, entiendo que lo importante es que las personas que ocupan puestos de poder y ostentan la capacidad para tomar decisiones, especialmente en el plano legislativo, tengan la mente abierta y, al menos, escuchen las propuestas de aquellos que de forma humilde y proactiva tratan de proponer soluciones (sin perjuicio, por supuesto, de que luego tengan la libertad de, en función de sus posiciones políticas de cada momento, hacerlas suyas o no).

De acuerdo con lo expuesto, debo afirmar que la motivación que subyace en la presente tesis doctoral no es otra que la de investigar qué posibilidades nos ofrecen las nuevas tecnologías, en concreto la Inteligencia Artificial (en adelante, IA), para mejorar el servicio de la Administración de Justicia. Y, en concreto, y ante la gran cantidad de funciones que esta otorga, cuyo análisis total devendría inabarcable, he decidido ceñir el objeto de mi investigación a determinar qué posibilidades y utilidades puede ofrecer la IA en el ámbito de la investigación penal, en concreto en el procedimiento de instrucción español, y cuáles

[3] Delgado, 2020, pág. 69.

podrían ser sus potenciales riesgos. Y esa, justamente, es la hipótesis sobre la que se construye este trabajo: ¿puede la IA resultar de utilidad en la tarea de investigar delitos en el proceso penal de instrucción español?.

Y es que, por un lado, la instrucción penal es un campo que despierta mucho interés en mí y que conozco razonablemente bien, habida cuenta de que durante los últimos años he servido como juez en un Juzgado de Instrucción; y, por otro lado, la IA es una tecnología que me atrae desde hace tiempo, ya que considero que tiene mucho que ofrecer y es "la gran desconocida". Y alguien se tenía que poner a investigar sobre tal cuestión que, sin duda, no resulta rentable para las empresas privadas que son las que suelen invertir recursos personales y materiales en explorar las posibilidades tecnológicas más novedosas, pero se antoja necesaria en el ámbito de la justicia penal.

Si bien es cierto que la realidad evoluciona tan rápido que para la ciencia jurídica y, en concreto, para el poder legislativo -que cuenta con grandes limitaciones burocráticas, de procedimiento y, por ende, temporales-, resulta imposible establecer regulaciones que vayan a la misma velocidad, lo cierto es que a futuro no puede permitirse que, aun a sabiendas de que existen herramientas a nuestra disposición que permitirían hacer mucho más eficiente, en todos los sentidos, un servicio público tan fundamental como es el de la Administración de Justicia, se obvie tal realidad (por falta de interés, por desconocimiento, por falta de tiempo, por falta de medios o simplemente por las dificultades que se dan en la pesada maquinaria de la Administración), y se permita que nuestro país se quede a la cola del mundo y condene al ciudadano, así, a conformarse con un servicio que (sabido por todos) es altamente deficiente y mejorable.

La introducción de las nuevas tecnologías en el ámbito de la Administración de Justicia, no obstante, no es algo que pueda hacerse de forma rápida, liviana y precipitada, puesto que hay numerosos derechos fundamentales que están en juego, por lo que en todo caso procede llevar a cabo un estudio y un análisis pormenorizado previo de los posibles beneficios y los potenciales riesgos que estas pueden entrañar con el fin de establecer una regulación garantista que, a su vez, no limite la evolución.

Y de ello, en nuestro Derecho Procesal Penal ya tenemos antecedentes (y, además, exitosos), habida cuenta de que el Título VIII del Libro II del Real Decreto de 14 de septiembre de 1882, aprobatorio de la Ley de Enjuiciamiento Criminal, ya fue modificado por la Ley Orgánica 13/2015, de 5 de octubre, de modificación de la Ley de Enjuiciamiento Criminal para el fortalecimiento de las garantías procesales y la regulación de las medidas de investigación tecnológica. Y es que, en virtud de tal reforma, se procedió a regular en nuestra Ley de Enjuiciamiento Criminal el uso de una serie de medidas de investigación tecnológica (a saber, la interceptación de las comunicaciones telefónicas y telemáticas, la captación y grabación de comunicaciones orales mediante la utilización de dispositivos electrónicos, la utilización de dispositivos técnicos de seguimiento, localización y captación de la imagen, el registro de dispositivos de almacenamiento masivo de información y los registros remotos sobre equipos informáticos) que, sin duda, han marcado un antes y un después en la forma de instruir las causas y, que, desde luego, han supuesto una revolución que ha reportado enormes beneficios en el ámbito de la investigación penal. Y es que tales medidas, no solo han implicado avances en la investigación criminal, sino que lo han hecho de forma garantista y respetuosa con los derechos fundamentales de los ciudadanos, al exigirse control judicial para su adopción, lo cual es encomiable y, desde luego, debe servir de base para la eventual futura regulación de la IA.

Mi misión y mi responsabilidad como juez, sin duda, ha de ser la de intentar destruir por fin ese estereotipo de "justicia lenta e ineficiente" que pesa sobre la Administración de Justicia española y que, en realidad, no hace justicia, valga la redundancia, ni representa a la enorme cantidad de grandísimos profesionales que hay detrás de toda la "vieja maquinaria" visible, que cada día se desgañitan para sacar adelante el papel de los juzgados con plena consciencia de que, detrás de cada expediente hay personas y, en muchas ocasiones, se hallan viviendo auténticos dramas que requieren de una solución rápida y de calidad, pero se ven muy limitados por la falta de medios existentes.

Y con ello no hablo de ir cambiando el Derecho Penal y el Derecho Procesal Penal al mismo ritmo que la tecnología, ni mucho menos (básicamente porque ello, como ya se ha dicho, es absolutamente imposible), sino de ir readaptando las nuevas herramientas disponibles a los principios y valores del mismo, para evitar justamente que su aplicación

sin pautas legales claras cause efectos nocivos. Y es que, nos guste o no, con falta de medios los seres humanos no podemos llegar a todo lo que se espera de nosotros y de la forma que nos gustaría, especialmente en el mundo "moderno" en el que vivimos, lo cual resulta muy frustrante y, además, en concreto, en ocasiones intolerable en el Estado del Bienestar en que nos hallamos.

Así, mi único objetivo con este trabajo es el de aportar mi "granito de arena" a la imprescindible evolución de la Administración de Justicia, que sin duda sé que está lejos, pero deseablemente cada vez más cerca. Y mi principal fin es conseguir la mejora del servicio otorgado por los órganos de instrucción a los ciudadanos mediante la introducción de una tecnología tan novedosa como la IA, con pleno respeto siempre, no obstante, a los derechos y libertades fundamentales y al resto del ordenamiento jurídico.

Ello, espero, además, que traspase los límites estrictamente jurídicos y genere efectos de mayor impacto social, habida cuenta de que un buen control y una buena gestión de la delincuencia van estrictamente relacionados con una mejora de la calidad de vida de los ciudadanos, por el aumento de su seguridad y, por ende, de su libertad en todos los sentidos.

No obstante lo anterior, conviene dejar claro desde el principio que este no es un trabajo sobre IA, sino sobre Derecho Penal y Derecho Procesal Penal. Y es que si bien, evidentemente, para entender los posibles beneficios y los potenciales riesgos que la IA puede entrañar en la investigación criminal he tenido que hacer una labor previa de investigación y comprensión de los conceptos básicos que la conforman, lo cierto es que me he centrado en profundizar sobre las cuestiones jurídicas que resultan aplicables, habida cuenta de que no soy ni ingeniera ni informática, soy jurista.

Una vez expuesto lo anterior, entiendo procedente exponer cuál ha sido el esquema y el orden de conceptos que he decidido seguir en el presente trabajo con el fin de facilitar su comprensión, dar una línea argumental clara y sistematizar ideas.

Así, en primer lugar, he considerado necesario dedicar un apartado inicial al estudio de una serie de conceptos y cuestiones generales relacionados con la IA, cuya comprensión entiendo indispensable para poder dar sentido a este trabajo. Y es que ante una temática

tan novedosa como la que trato, pienso que es fundamental dejar bien sentadas las bases conceptuales y los asuntos jurídicos subyacentes desde el principio para poder así permitir al lector adentrarse en la posterior parte más específica con las ideas más claras e, incluso, con una incipiente opinión formada.

En relación con ello, pues, he decidido hacer especial referencia al complejo y difuso concepto de IA; he entendido oportuno hacer un recorrido por la historia de tal aparentemente nueva tecnología, con el fin de hacer ver que no es algo tan reciente como generalmente se cree; y he considerado necesario clasificarla en función de varios de sus rasgos y características, habida cuenta de que existen múltiples tipos de sistemas que emplean tal tecnología y no todos funcionan del mismo modo.

Asimismo, he pensado que, desde el punto de vista jurídico, podría resultar interesante hacer un análisis de los posibles beneficios y los potenciales riesgos que, de forma general, puede entrañar la IA para los derechos y las libertades de los ciudadanos, para lo cual he considerado oportuno, por un lado, hacer una serie de reflexiones sobre los conceptos de ética y moral; por otro lado, hacer una recopilación de todas aquellas iniciativas y propuestas privadas y públicas relativas a la regulación de tal tecnología; he decidido plasmar mi punto de vista sobre cuáles son aquellas decisiones que deberían tomarse para garantizar un uso mundial de la IA con fines beneficiosos para el ser humano, limitando su utilización para evitar así una desnaturalización de la especie; y, finalmente, he decidido concluir con una propuesta relativa a los principios básicos sobre IA que considero que deberían regir en todo caso para conseguir los fines expuestos y evitar vulneraciones masivas de derechos y libertades.

Tras ello, en segundo lugar, de forma más específica y profunda, he decidido pasar a analizar lo que resulta ser el *alma mater* de la presente tesis doctoral: las aplicaciones de la IA en el ámbito de la investigación criminal y, en concreto, en el proceso penal de instrucción español.

Respecto de ello, con carácter previo, he considerado necesario hacer una breve referencia al mencionado proceso penal de instrucción español y a la posible introducción de la IA en el mismo como herramienta de investigación, con especial mención a la regulación de

medios de investigación tecnológicos ya presente en el Real Decreto de 14 de septiembre de 1882, aprobatorio de la Ley de Enjuiciamiento Criminal. Y, posteriormente, he entendido oportuno adentrarme en lo que es el grueso y, sin duda, el tema estrella de este trabajo, a saber: las herramientas de IA para investigar delitos.

En relación con ello, he decidido hacer una clasificación en tres bloques distintos, por las diferencias concurrentes entre ellas, que considero que puede ayudar a entender mejor los conceptos y las posibles utilidades y potenciales riesgos que entrañan cada una.

Así, por un lado, he entendido oportuno proceder a analizar las herramientas de IA de predicción y evaluación de riesgos, es decir, aquellos sistemas que emplean tal tecnología para predecir eventos futuros y valorar la existencia de posibles y potenciales peligros o riesgos venideros mediante el examen de ingentes cantidades de datos históricos. Y, a la vista de que el uso de este tipo de instrumentos está ampliamente extendido en el ámbito policial (más que en el judicial, de hecho), he decidido hacer referencia al mismo, para lo que he entendido importante establecer una distinción entre los sistemas de mapeo predictivo; los sistemas de identificación predictiva; y los sistemas de evaluación de riesgos individuales, con especial alusión a su concepto, regulación y potenciales beneficios y riesgos. Y, por supuesto, también he decidido analizar la utilización de tales herramientas en el ámbito judicial, como sistemas de justicia predictiva, con examen de su regulación y de las posibles utilidades y riesgos jurídicos que pueden entrañar, haciendo especial mención al sistema estadounidense COMPAS.

Por otro lado, he decidido proceder a analizar las herramientas de investigación criminal propiamente dichas, es decir, aquellos sistemas que emplean tal tecnología y son utilizados por las autoridades policiales, fiscales y judiciales para esclarecer las circunstancias que rodean a los hechos presuntamente delictivos e identificar a sus autores.[4]

Y es que, si bien la gran mayoría de tales sistemas no fueron originariamente creados con la finalidad de ser destinados a la investigación criminal, lo cierto es que todos ellos cuentan con unas características y unas aptitudes que, sin duda, los hacen idóneos para lograr tal

[4] En relación con lo dispuesto en el artículo 299 de la LECrim.

objetivo. En tal sentido, una buena parte de mi labor de investigación consiste, justamente, en rastrear y analizar las múltiples y distintas herramientas existentes en el ámbito de la IA y hacer una cuidada y específica selección de aquellas que considere más adecuadas y aptas para ser puestas a disposición de las autoridades policiales, fiscales y judiciales con el fin de asistirles en la investigación penal.

Con el fin de sistematizar el estudio de este tipo de herramientas de IA, he entendido oportuno, además, establecer una distinción entre ellas, en función de la tecnología empleada.

Así, he decidido analizar en más profundidad, en primer lugar, las herramientas de IA que emplean datos biométricos, entre las que considero importante estudiar las de reconocimiento facial, las de reconocimiento de voz, las de reconocimiento de emociones, las de reconocimiento de huellas dactilares, las de reconocimiento de ADN, y las de reconocimiento de firma y escritura; en segundo lugar, las herramientas que emplean técnicas de Procesamiento del Lenguaje Natural (PLN), entre las que considero importante analizar los *chatbots* y otros sistemas, a saber: herramientas para detectar denuncias falsas, herramientas para detectar y, en su caso, moderar contenido *online*, herramientas para analizar documentos, herramientas de traducción simultánea, y herramientas de transcripción automática; en tercer lugar, las herramientas que emplean técnicas de Visión Artificial o *Computer Vision*, entre las que entiendo relevante analizar las de análisis de imágenes, las de lectura de matrículas y las de detección de documentos falsos; y, finalmente, las herramientas que emplean otras tecnologías, entre las que considero interesante examinar las de detección de estafas y fraudes digitales y las de detección de disparos.

Respecto de todas ellas considero fundamental abordar no solo su concepto, sino también sus posibles utilidades en la instrucción de las causas, que desde luego son en ocasiones inimaginables; su regulación, vigente y/o en proyecto, habida cuenta de que su uso debe quedar limitado por el marco jurídico aplicable en cada caso; y sus potenciales riesgos jurídicos, que no son ni pocos ni simples, y todo ello con el objetivo de otorgar una visión completa acerca de las mismas y proporcionar los elementos necesarios para poder determinar su eventual viabilidad jurídica.

Tras ello, una vez llevado a cabo el análisis de las mencionadas herramientas de IA de investigación penal, he entendido necesario hacer especial referencia al fenómeno de las vigilancias masivas, es decir, al control indiscriminado de los movimientos físicos y virtuales de un elevado número de personas -en ocasiones, incluso, de países enteros- que en buena parte se lleva a cabo mediante algunos de los sistemas antedichos. Y es que en los últimos años, numerosas organizaciones *pro* derechos humanos han ido denunciando de forma reiterada la práctica de tales vigilancias de forma indistinta y arbitraria por parte de algunos gobiernos, sin causa de justificación real alguna, siendo el de China el más interesante y llamativo de analizar.

Y, finalmente, he considerado conveniente hacer mención, de forma breve, a las herramientas de IA de tramitación de las causas penales (en concreto, en la fase de instrucción), puesto que la eficiencia en la gestión y la tramitación de los expedientes resulta imprescindible para poder conseguir dar un buen servicio, de calidad, al ciudadano. No obstante, y si bien es un terreno muy interesante de explorar, lo cierto es que la presente tesis doctoral debe quedar limitada a las posibles utilidades de la IA en el ámbito estrictamente jurídico y de contenido, habida cuenta de la necesidad de restringir la extensión y la temática.

En la actualidad, tal y como se pondrá de manifiesto a lo largo del presente trabajo, la regulación de la IA en España y en la Unión Europea (en adelante, UE) es prácticamente inexistente, y no solo respecto de su uso en el ámbito de la investigación criminal, sino en general, lo cual es para lamentar.

No obstante, por un lado, es importante determinar cuáles son las implicaciones que podría tener la vigente legislación en materia de protección de datos personales en la utilización de herramientas de IA en el proceso de instrucción español y, en concreto, cuál es el alcance de lo dispuesto en la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales. Para ello, se reputa interesante llevar a cabo un análisis transversal del distinto alcance e impacto que puede tener tal

legislación respecto de cada una de las herramientas de IA que se analizarán en este trabajo, habida cuenta de que esta puede ser mayor o menor según su naturaleza y su contenido.

Por otro lado, es interesante poner el foco en la futura legislación sobre IA que tiene entre manos la UE y que, por el momento, únicamente cuenta con el valor de propuesta y consta plasmada en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, a la espera de publicación del texto definitivo tras la aprobación del Consejo y del Parlamento de la UE, lo cual se prevé para el año 2022.

Para ello, pues, he considerado oportuno efectuar un estudio pormenorizado de sus posibles implicaciones en cada uno de los tipos de herramientas de IA que se van a analizar. Sin perjuicio, desde luego, de lo que pueda finalmente resultar cuando se publique el texto definitivo, lo cual justificará, sin duda, el inicio de una investigación *post* doctoral.

Y es que la IA es una tecnología que, como ya se ha avanzado, está en constante evolución, y no solo desde un punto de vista técnico, sino también jurídico, puesto que el actual interés por sus posibles beneficios y, sobre todo, la preocupación por sus potenciales riesgos, no hace más que agitar de forma incesante el panorama tecnológico y legislativo.

Y ello no es algo que se circunscriba únicamente a la UE o a España, ni muchísimo menos, ello es algo que traspasa fronteras y que no entiende de jurisdicciones o soberanías, puesto que allá donde surge la posibilidad de aplicar IA se inicia automáticamente una revolución tecnológica y jurídica.

En relación con ello, considero absolutamente fundamental que el estudio de las posibles implicaciones jurídicas que la IA puede entrañar se lleve a cabo siempre desde una perspectiva internacional, habida cuenta de que un estudio más limitado, meramente nacional, por ejemplo, podría resultar poco profundo y podría llevar a sacar conclusiones erróneas, puesto que la eclosión de tal tecnología es un fenómeno global y complejo, que puede tener implicaciones para toda la humanidad y de tal forma debe ser tratado.

Con tal intención, desde el momento en que decidí iniciar el presente trabajo tuve claro que era necesario realizar una colaboración con alguna universidad extranjera para hacer una estancia de investigación fuera de España a fin de poder instruirme sobre el uso y las implicaciones jurídicas que la IA está teniendo en otros países, especialmente en Estados Unidos, donde sin duda llevan ventaja en el uso de tal tecnología, y en Argentina, donde cuentan con pioneras e interesantes iniciativas en el ámbito del uso de la IA en la investigación penal.

Así, para la elaboración de este trabajo he realizado una estancia de investigación en la Facultad de Derecho de la Universidad de Georgetown (Washington DC, Estados Unidos -en adelante, EEUU-), donde he podido observar cómo se enfoca el uso de la IA en tal país, que está a la carrera con China, Rusia e Israel por ostentar la hegemonía en tal materia, y cómo se tratan los posibles riesgos que esta puede entrañar para los derechos y libertades de los ciudadanos, especialmente en el ámbito de la justicia penal.

Asimismo, he llevado a cabo una colaboración con la Universidad Austral de Buenos Aires (Argentina), que me ha permitido conocer de primera mano el contenido de los proyectos que, en especial, desde el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires, se están llevando a cabo en relación con el uso de la IA en la investigación penal, especialmente en el ámbito de la tramitación de las causas, lo que sin duda resulta un espejo donde mirarse desde la UE y desde España.

En cualquier caso, tras efectuar un amplio y complejo trabajo de investigación, dentro y fuera de España, y a pesar de que uno siempre tiene la sensación de poder llegar "más allá", lo que tengo claro es que la tesis doctoral que presento es fruto de una motivación de cambio profunda y honesta, que surge de mi condición de servidora pública y que, espero, en el futuro pueda resultar de utilidad para mejorar el servicio que la Administración de Justicia ofrece a los ciudadanos, sin que ello suponga una limitación arbitraria de sus derechos y libertades y, asimismo, pueda conllevar un aumento de la calidad de vida de nuestras sociedades como consecuencia de una más eficiente y exitosa gestión de la realidad criminal.

Lograr tal objetivo de forma real y efectiva, no obstante, por suerte o por desgracia, desde luego, no depende de mí, sino que es responsabilidad del poder legislativo que ostente el poder en cada momento, ya que tal y como asegura Steve M. Bellovin, Profesor de informática de la Universidad de Columbia (Nueva York, EEUU): "*La toma de decisiones algorítmica representa una forma neoliberal de creación de políticas* (…)."[5]

De todas formas, entiendo que es tarea y compromiso de todos auxiliar en la medida de lo posible a que las decisiones políticas sean tomadas contando con la mejor y la mayor información posible, y espero que, algún día, mi aportación y mi trabajo puedan valer para ello.

Empezamos.

[5] Waldman, 2019, pág. 12.

1BIS.- SUMMARY

Already in the 19th century, Mariano José de Larra (1809-1837), one of the greats of Spanish political romanticism, satirized in his well-known article entitled "Come back tomorrow" [6] the reality of the functioning of the services in Spain and, specifically, of the Public Administration.

In this story he exposed with great sarcasm how a foreigner arrived in our country determined to make an investment, with the idea of leaving all the paperwork finished in fifteen days, and ended up in despair after months due to the slowness and inefficiency of all those who should provide collaboration, especially administrative.

In this sense, I consider it very illustrative to bring up some of the brilliant passages of the aforementioned article which, without a doubt unfortunately, not only reflect the reality of the Spanish Administration of the 19th century, but also that of the current one:

"*On returning from the report, it was realized in the section of our blessed office that the said file did not correspond to that branch; it was necessary to rectify this small error; it was passed to the branch, establishment and corresponding desk, and here we are, after three months, always on the tail of our file, like a ferret looking for a rabbit, and without being able to get it out of the ferret's box either dead or alive. It was the case when we arrived here that the file left the first establishment and never arrived at the other.*
-From here it was sent with the date of so many-they said in one.
-Nothing has arrived here- said another.
-"*Voto va*"*! I said to Monsieur Sans-délai. Do you know that our dossier has been left in the air like the soul of Garibay, and that it must now be perched like a dove on some roof of this busy town? Another one had to be made. Back to pawns! Back to haste! What a delirium!*
-It is indispensable- said the officer in a bell-like voice, that these things go through their regular channels. (…)

[6] De Larra, 1833.

After six months I have only succeeded in being told everywhere every day: "Come back tomorrow", and when this blessed "tomorrow" finally arrives, they tell us roundly that "no"? (…)"

It is well known that, at present, the Administration of Justice is the worst valued Public Administration by the Spanish population, since just one in five Spaniards approves its operation, according to a survey published by the CIS in 2020.[7] And this, in my view, is unacceptable.

Unacceptable because it is clear that something is wrong. Although most of the professionals of the Administration of Justice work hard to ensure that the procedures go ahead in the best possible way, the truth is that justice does not work. It is slow, it is inefficient and, therefore, it does not offer a quality service.

And this, of course, is not a problem suffered only by the citizens who have to wait years, often with anguish, for their lawsuits to be resolved. It is a problem that also affects the people who work in the Administration, who in most cases are completely overwhelmed by the workload they have, which is humanly impossible to digest. And, obviously, in the face of such a problem, we all lose.

This is undoubtedly the circumstance that, after the first years of my professional career, prompted me to think about possible solutions. Solutions that I saw as indispensable to be able to continue to dedicate myself to my profession, the judiciary, since the current working conditions are not (or should not be) sustainable, especially in a world where we have at our disposal a myriad of new technologies that can be extremely useful to improve the service provided by the Administration. And I understand that, in my position as a public servant, this is not a mere "survival" or "goodwill" initiative, it is a real responsibility.

However, the truth is that, unfortunately, not all members of the Administration of Justice, for one reason or another, have the same attitude to encourage and accept the introduction

[7] See Castro, 2020.

of changes, especially technological ones. And, in relation to this, I think it is appropriate to bring up what Carlos Delgado Romero, Head of the Forensic Acoustics Laboratory of the General Commissariat of Scientific Police of the National Police Corps, who paraphrasing the American philosopher Thomas Khun, said: "*I do what I can and the best I can with the resources I have; let's see if now that I have everything organized someone comes and tells me that I have to change the way I work*." [8]

And such an attitude, of course, is not only a limitation for evolution, but also a real delay, since such individuals not only do not contribute with new ideas to improve the service provided by the Administration of Justice, but also show their frontal opposition to any novelty that may involve changes, and therefore hinder the path of progress.

In any case, "it never rains to everyone's taste" and that, without doubt, is something that has to be accepted, since it is inherent in human diversity. However, I understand that the important thing is that people who occupy positions of power and have the capacity to make decisions, especially at the legislative level, have an open mind and at least listen to the proposals of those who proactively try to propose solutions (without prejudice, of course, that they are then free, depending on their political positions at any given time, to make them their own or not).

In accordance with the above, I must state that the motivation underlying this doctoral thesis is none other than to investigate the possibilities offered by new technologies, specifically AI, to improve the service of the Administration of Justice. And, in particular, and given the large number of functions that it provides, whose total analysis would become unmanageable, I have decided to limit the object of my research to determine what possibilities and utilities AI can offer in the field of criminal investigation, specifically in the Spanish investigative procedure. And that, precisely, is the hypothesis on which this paper is built: can AI be useful in the task of investigating crimes in the Spanish investigative process?

[8] Delgado, 2020, p. 69.

On the one hand, criminal investigation is a field that arouses great interest in me and that I know reasonably well, given that for the last five years I have served as an investigation judge; and, on the other hand, AI is a technology that has attracted me for years, as I consider that it has much to offer and it is "the great unknown". And so, someone had to start researching on such an issue that, undoubtedly, is not profitable for private companies that are those who usually invest personal and material resources in exploring the latest technological possibilities, but it seems necessary in the field of criminal justice.

Although it is true that reality evolves so fast that it is impossible for legal science and, specifically, for the legislative power -which has great bureaucratic, procedural and, therefore, temporal limitations- to establish regulations that go at the same speed, the truth is that in the future it cannot be allowed that, even knowing that there are tools at our disposal that would allow us to make such a fundamental public service as the Administration of Justice much more efficient, in all senses of the word, we ignore this reality (due to lack of interest, lack of knowledge, lack of time, lack of means or simply due to difficulties in the Administration of Justice), and we relegate our country to remain at the bottom of the world and condemn the citizen, thus, to settle for a service that (as we all know) is highly deficient and can be improved.

The introduction of new technologies in the field of the Administration of Justice, however, is not something that can be done quickly, lightly and hastily, since there are many fundamental rights at stake, so that in any case it is appropriate to carry out a study and a detailed prior analysis of the possible benefits and potential risks that these may entail in order to establish a guaranteeing regulation that, in turn, does not limit the evolution.

And of this, in our Criminal Procedural Law we already have precedents (and, moreover, successful), given that Title VIII of Book II of the Royal Decree of September 14th 1882, approving the Criminal Procedure Act, was already amended by Organic Law 13/2015, of October 5th, amending the Criminal Procedure Act for the strengthening of procedural guarantees and the regulation of technological investigative measures. And, by virtue of such reform, our Criminal Procedure Act proceeded to regulate the use of a series of technological investigative measures (namely, the interception of telephone and telematic communications, the capture and recording of oral communications through the use of

electronic devices, the use of technical tracking, location and image capture devices; and the registry of mass storage devices and remote registries on computer equipment). These measures have undoubtedly marked a before and after in the way of investigating cases and, of course, have been a revolution that has brought enormous benefits in the field of criminal investigation. And the fact is that such measures have not only implied advances in criminal investigation, but have done so in a way that guarantees and respects the fundamental rights of citizens, by requiring judicial control for their adoption, which is commendable and, of course, should serve as a basis for the eventual future regulation of AI.

My mission and my responsibility as a judge, undoubtedly, must be to try to finally destroy the stereotype of "slow and inefficient justice" that weighs on the Spanish Administration of Justice and that, in reality, does not do justice, it is worth the redundancy, nor does it represent the enormous amount of great professionals that are behind all the visible "old machinery", who every day strive to move the role of the courts forward with full awareness that, behind each file there are people and, on many occasions, are living real dramas that require a quick and quality solution, but are severely limited by the lack of existing means.

And by this I am not talking about changing Criminal Law and Criminal Procedural Law at the same pace as technology, far from it (basically because this, as has already been said, is absolutely impossible), but to readapt the new tools available to the principles and values of the same, in order to prevent their application without clear legal guidelines from causing harmful effects. The fact is that, whether we like it or not, with a lack of means we human beings cannot achieve everything that is expected of us and in the way we would like to, especially in the "modern" world in which we live, which is very frustrating and, moreover, in particular, sometimes intolerable in the Welfare State in which we find ourselves.

Thus, my only objective with this work is to contribute my "grain of sand" to the essential evolution of the Administration of Justice, which I know is undoubtedly far away, but desirably getting closer. And my main aim is to achieve the improvement of the service provided by the investigative bodies to citizens through the introduction of such a novel

technology as AI, always fully respecting, however, the fundamental rights and freedoms and the rest of the legal system.

I also hope that this will go beyond strictly legal limits and have a greater social impact, given that good crime control and good crime management are strictly linked to an improvement in the quality of life of the citizens, by increasing their security and, therefore, their freedom in all senses of the word.

Notwithstanding the above, it should be made clear from the outset that this is not a paper on AI, but on Criminal Law and Criminal Procedural Law. And although, obviously, in order to understand the possible benefits and potential risks that AI may entail in criminal investigation, I have had to do some previous research and understanding of the basic concepts that surround it, the truth is that I have focused on delving into the legal issues that are applicable, given that I am neither an engineer nor a computer scientist, I am a lawyer.

Having stated the above, I consider it appropriate to explain the scheme and the order of concepts that I have decided to follow in this work in order to facilitate its understanding, to give a clear line of argument and to systematize ideas.

Thus, first of all, I have considered it necessary to dedicate an initial section to the study of a series of concepts and general issues related to AI, whose understanding is essential in order to give meaning to this work. In the case of a subject as novel as the one I am dealing with, I think it is essential to establish the conceptual foundations and the underlying legal issues from the beginning in order to allow the reader to enter the subsequent more specific part with clear ideas and, even, with an incipiently formed opinion.

In relation to this, then, I have decided to make special reference to the complex and diffuse concept of AI; I have seen fit to make a tour through the history of this apparently new technology, in order to show that it is not something as recent as is generally believed; and I have considered it necessary to classify it according to several of its features and characteristics, given that there are multiple types of systems that use such technology and not all work in the same way.

I also thought that, from a legal point of view, it might be interesting to make an analysis of the possible benefits and potential risks that, in general, AI may entail for the rights and freedoms of citizens, for which I have considered it appropriate, on the one hand, to make a series of reflections on the concepts of ethics and morality; on the other hand, to make a compilation of all those initiatives and private and public proposals relating to the regulation of such technology; I have decided to express my point of view on the decisions that should be taken to guarantee a worldwide use of AI with beneficial purposes for human beings, limiting its use in order to avoid a denaturalization of the species; and, finally, I have decided to conclude with a proposal regarding the basic principles on AI that I believe should govern in any case in order to achieve the aforementioned purposes and avoid massive violations of rights and freedoms.

After that, secondly, in a more specific and deeper way, I have decided to analyze what turns out to be the *alma mater* of the present doctoral thesis: the applications of AI in the field of criminal investigation and, specifically, in the Spanish investigative process.

In this regard, I have previously considered it necessary to make a brief reference to the aforementioned Spanish investigative process and the possible introduction of AI in it as an investigative tool, with special mention to the regulation of technological means of investigation already present in the Royal Decree of 14 September 1882, approving the Criminal Procedure Act. Subsequently, I have seen fit to enter into what is the bulk and, undoubtedly, the focal topic of this work, namely: AI tools for investigating crimes.

In relation to them, I have decided to make a classification in three different blocks, due to the concurrent differences between them, which I believe can help to better understand the concepts and the possible utilities and potential risks involved in each one.

Thus, on the one hand, I have considered it appropriate to proceed to analyze AI risk assessment tools, i.e., those systems that use such technology to predict future events and assess the existence of possible and potential dangers or risks to come by examining huge amounts of historical data, being able to estimate the probability of something happening. And, in view of the fact that the use of this type of tools is widely extended in the police

field (more than in the judicial field, in fact), I have decided to make reference to it, for what I have understood important to establish a distinction between predictive mapping systems; predictive identification systems; and individual risk assessment systems, with special reference to their concept, regulation and potential risks and benefits. And, of course, I have also decided to analyze the use of such tools in the judicial field, as predictive justice systems, with an examination of their regulation and the possible utilities and legal risks that may arise, with special mention of the American COMPAS system.

On the other hand, I have decided to proceed to analyze the AI criminal investigation tools themselves, that is, those systems that employ such technology and are used by the police, prosecutors and judicial authorities to clarify the circumstances surrounding the alleged criminal acts and identify their perpetrators.[9]

Although the vast majority of these systems were not originally created with the aim of being used for criminal investigation, the fact is that all of them have characteristics and skills that undoubtedly make them suitable to achieve that goal. In this sense, a large part of my research work will consist, precisely, in tracking and analyzing the many different existing tools in the field of AI and make a careful and specific selection of those that I consider most appropriate and suitable to be made available to the police, prosecutors and judicial authorities in order to assist them in criminal investigation.

In order to systematize the study of this type of AI tools, I have considered it appropriate, in addition, to establish a distinction between them, depending on the technology used.

Thus, I will proceed to analyze in more depth, first, AI tools that employ biometric data, among which I consider it important to study facial recognition tools, voice recognition tools, emotion recognition tools, fingerprint recognition tools, DNA recognition tools, and signature and handwriting recognition tools; second, tools that employ Natural Language Processing (NLP) techniques, among which I consider it important to analyze chatbots and other systems, viz: tools to detect false reports, tools to detect and, where appropriate, moderate online content, tools to analyze documents, simultaneous translation tools, and

[9] In relation to the provisions of Article 299 of the Spanish Criminal Procedure Act.

automatic transcription tools; thirdly, tools that employ Computer Vision techniques, among which I consider it relevant to analyze image analysis tools, license plate reading tools and false document detection tools; and, finally, tools that employ other technologies, among which I consider it interesting to examine those for detecting scams and digital fraud and those for detecting gunshots.

With regard to all of them, I consider it essential to address not only their concept, but also their possible uses in the investigation of cases, which are of course sometimes unimaginable; their regulation, in force and/or in the pipeline, given that their use must be limited by the legal framework applicable in each case; and their legal risks, which are neither few nor simple, all with the aim of providing a complete view of them and providing the necessary elements to determine their possible legal viability.

After this, having carried out the analysis of the aforementioned AI tools for criminal investigation, I have found it necessary to make special reference to the phenomenon of mass surveillance, that is, the indiscriminate control of the physical and virtual movements of a large number of people -sometimes even entire countries- which is largely carried out through some of the aforementioned systems. In recent years, numerous human rights organizations have repeatedly denounced the practice of such surveillance by some governments in an indiscriminate and arbitrary manner, without any real justification, with China's being the most interesting and striking case to analyze.

Finally, I have considered it appropriate to mention, briefly, AI tools for the processing of criminal cases (specifically, in the pre-trial phase), since efficiency in the management and processing of files is essential in order to provide a good quality service to citizens. However, and although it is a very interesting field to explore, the truth is that this doctoral thesis should be limited to the possible uses of AI in the strictly legal and content area, given the need to restrict the extension and subject matter.

At present, as will be shown throughout this paper, the regulation of AI in Spain and in the European Union (hereinafter, EU) is practically non-existent, and not only with respect to its use in the field of criminal investigation, but in general, which is regrettable.

However, on the one hand, it is important to determine the implications that the current legislation on personal data protection could have on the use of AI tools in the Spanish investigative process and, specifically, the scope of the provisions of Organic Law 7/2021, of May 26th, on the protection of personal data processed for the purposes of prevention, detection, investigation and prosecution of criminal offences and the execution of criminal penalties. To this end, it is considered interesting to carry out a cross-sectional analysis of the different scope and impact that such legislation may have with respect to each of the AI tools that will be analyzed in this work, given that this may be greater or lesser depending on their nature and content.

On the other hand, it is interesting to focus on the future legislation of AI that the EU is preparing and that, for the moment, only has the value of a proposal and is embodied in the Proposal for a Regulation of the European Parliament and of the Council establishing Harmonized Rules in the field of Artificial Intelligence (Artificial Intelligence Act) and amending certain legislative acts of the Union, pending the publication of the final text after the approval of the Council and the Parliament of the EU, which is expected for 2022.

To this end, therefore, I have considered it appropriate to carry out a detailed study of their possible implications for each of the types of AI tools to be analysed. Without prejudice, of course, to what may finally result when the final text is published, which will undoubtedly justify the start of a postdoctoral research.

AI is a technology that, as has already been mentioned, is constantly evolving, and not only from a technical point of view, but also from a legal one, since the current interest in its possible benefits and, above all, the concern about its potential risks, stirs the technological and legislative landscape incessantly.

Moreover, this not something that is limited only to the EU or Spain, far from it, it is something that crosses borders and does not understand about jurisdictions or sovereignties, since wherever the possibility of applying AI arises, a technological and legal revolution automatically begins.

In relation to this, I consider it absolutely fundamental that the study of the possible legal implications that AI may entail should always be carried out from an international perspective, given that a more limited, merely national study, for example, could be shallow and could lead to erroneous conclusions, since the emergence of such technology is a global and complex phenomenon, which may have implications for the whole humanity and must be dealt with in such a way.

With this intention, from the moment I decided to start this work it was clear to me that it was necessary to collaborate with a foreign university to do a research stay outside Spain in order to learn about the use and legal implications that AI is having in other countries, especially in the United States, where they are certainly ahead in the use of such technology, and in Argentina, where they have pioneering and interesting initiatives in the field of the use of AI in criminal investigation.

Thus, for the preparation of this paper I have carried out a research stay at Georgetown University Law School (Washington DC, USA), where I have been able to observe how the use of AI is approached in that country, which is in the race with China, Russia and Israel to hold the hegemony in this field, and how the possible risks that this may entail for the rights and freedoms of citizens, especially in the field of criminal justice, are dealt with.

Likewise, I have carried out a collaboration with the Austral University of Buenos Aires (Argentina), which has allowed me to know first-hand the content of the projects that, especially from the Public Prosecutor's Office of the Autonomous City of Buenos Aires, are being carried out in relation to the use of AI in criminal investigation, especially with regard to the processing of cases, which is undoubtedly a mirror to look at from the EU and from Spain.

In any case, after carrying out an extensive and complex research work, inside and outside Spain, and despite the fact that one always has the feeling of wanting to go "further", what is clear to me is that the doctoral thesis that I am presenting is the result of a motivation for profound, honest and ambitious change, which arises from my condition as a public servant and which, I hope, in the future may be useful for improving the service that the Administration of Justice offers to citizens, without this entailing an arbitrary limitation of

their rights and freedoms and, likewise, may lead to an increase in the quality of life of our societies as a result of a more efficient and successful management of the criminal reality.

Achieving such an objective in a real and effective way, however, fortunately or unfortunately, of course, does not depend on me, but it is the responsibility of the legislative power that holds the power at each moment, since as Steve M. Bellovin, Professor of Computer Science at Columbia University (NY, USA) states: "*Algorithmic decision making represents a neoliberal form of policy making* (…)."[10]

In any case, I understand that it is everyone's task and commitment to help, as far as possible, to ensure that political decisions are taken on the basis of the best and most complete information possible, and I hope that, one day, my contribution and my work will help to achieve this.

Let's get started.

[10] Waldman, 2019, pág. 12.

2.- IA: CUESTIONES GENERALES

2.1. ¿QUÉ ES LA IA?

Winston Churchill, en el discurso pronunciado el 6 de septiembre de 1943 en la Universidad de Harvard, tras recibir el título de Doctor *honoris causa*, predijo con gran acierto: "*Los imperios del futuro son los imperios de la mente*"[11]. No obstante, no resulta osado adivinar que dicho sabio y carismático político jamás pudo llegar a imaginar qué dimensión llegarían a tomar sus palabras, ya que tan solo unos años más tarde el término "mente" al que hacía referencia dejaría de ser monopolio de los humanos para pasar a ser compartido con las máquinas.

Con carácter previo a adentrarme en la investigación que otorga título a la presente tesis doctoral, resulta absolutamente imprescindible abordar de forma breve la cuestión conceptual del fenómeno que subyace a lo largo de la misma: la IA.

Dar una definición del concepto de IA, al contrario de lo que podría parecer, es una de las cuestiones más complejas y desafiantes a las que se enfrentan los estudiosos de esta tecnología, ya que resulta muy difícil delimitar, sin caer en la obsolescencia, un concepto que evoluciona tan deprisa[12].

Para poder entender qué es la IA algunos autores consideran que primero debe comprenderse qué es la inteligencia humana[13], lo cual resulta ciertamente complicado, habida cuenta de los múltiples matices que esta alberga, tal y como recoge, entre otros, Howard Gardner en su reconocida obra "*Frames of Mind: The Theory of multiple intelligences*".[14]

En mi opinión, la comparación directa entre ambas inteligencias no es el mejor criterio para poder llegar a entender qué es la IA, puesto que, por un lado, "*las máquinas pueden realizar*

[11] Churchill, 1943
[12] Necati, Cervera, Cuatrecasas, Keser, Atabey & otros, 2020, pág. 6.
[13] Entre otros, Jeff Hawkins y Sandra Blakeslee.
[14] Gardner, 1983.

tareas que las personas no pueden en absoluto"[15] y, por otro lado, los humanos contamos con una sensibilidad, un sentido común y una capacidad de improvisación que las máquinas, al menos por el momento, no llegan a alcanzar (a pesar de que a lo largo de la historia se han ido desmontando, sin piedad, las previsiones humanas sobre los hitos que las máquinas nunca podrían llegar a alcanzar). Así, la comparación entre la inteligencia humana y la artificial implicaría la equiparación de dos ideas con contenido no idéntico que podría inducir a error y/o confusión, por lo que desde mi punto de vista la mejor opción es otorgar a la IA una definición autónoma.

Recientemente, el High-Level Expert Group on sustainable finance (HLEG) de la Comisión Europea definió la IA como aquellos "*sistemas de software (y posiblemente también de hardware) diseñados por humanos que, ante un objetivo complejo, actúan en la dimensión física o digital: percibiendo su entorno, a través de la adquisición e interpretación de datos estructurados o no estructurados, razonando sobre el conocimiento, procesando la información derivada de estos datos y decidiendo las mejores acciones para lograr el objetivo dado. Los sistemas de IA pueden usar reglas simbólicas o aprender un modelo numérico, y también pueden adaptar su comportamiento al analizar cómo el medio ambiente se ve afectado por sus acciones previas*".[16]

Y, por su parte, en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, publicada el 21 de abril de 2021, se definen los sistemas de IA en su artículo 3.1 como: "*el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I* (*a saber, Estrategias de aprendizaje automático, incluidos el aprendizaje supervisado, el no supervisado y el realizado por refuerzo, que emplean una amplia variedad de métodos, entre ellos el aprendizaje profundo. Estrategias basadas en la lógica y el conocimiento, especialmente la representación del conocimiento, la programación (lógica) inductiva, las bases de conocimiento, los motores de inferencia y deducción, los sistemas expertos y de razonamiento (simbólico). Estrategias estadísticas, estimación bayesiana, métodos de búsqueda y optimización.*) *y que puede, para un*

[15] Kaplan, 2017, pág. 4.
[16] Comisión Europea, 2020.

conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa."

A mi entender, no obstante, una de las definiciones de la IA más completas que existen hoy en día es aquella que establece que la "*La IA es aquella actividad dedicada a hacer que las máquinas sean inteligentes, y la inteligencia es esa cualidad que permite a una entidad funcionar de manera adecuada y con previsión en su entorno*."[17], resultando la esencia de la IA (y, en mi opinión, de la inteligencia, en general) la capacidad para hacer generalizaciones adecuadas de un modo oportuno, basándose en datos limitados.

Y en tal sentido, resulta interesante hacer especial mención a lo dispuesto por el Profesor Juan G. Corvalán, Director del Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires (Argentina), que aseveró: "*En todos los casos, las técnicas de IA se basan en detectar y reconocer patrones de información en los datos. Esto se logra a partir de combinar ordenadores, internet, algoritmos y lenguajes de programación para resolver problemas o tomar decisiones que antes solo podían ser realizadas por nuestras capacidades cognitivas*".[18]

Así, el empleo de la IA no tiene otra finalidad que la de llevar a cabo un comportamiento inteligente a través de las máquinas, lo que bajo mi punto de vista implica la necesidad de que estas observen, analicen, aprendan, interpreten y tomen decisiones por sí mismas, de modo que aporten alguna información o solución nueva al usuario.

Y es que si bien es cierto que hay una línea muy fina entre todo aquello que es IA y todo lo que no, hoy en día hay una cierta tendencia a calificar como tal todos los sistemas informáticos que escapan de nuestro conocimiento más básico, lo cual no siempre es ajustado a la realidad, puesto que en muchas ocasiones estos emplean meros procesos mecánicos que no analizan, interpretan, sacan conclusiones y toman decisiones por sí mismos, *conditio sine qua non*, en mi opinión, para poder afirmar que operan con IA.

[17] Nilsson, 2010.
[18] Dupuy & Corvalán, 2021, pág.9.

De acuerdo con lo expuesto, a lo largo de este trabajo cada vez que haga referencia a la IA, estaré haciendo alusión al concepto antedicho, si bien hay que decir que en ocasiones el encuadre de según qué sistemas en tal definición será más clara y evidente que en otras, puesto que no debemos olvidar que no se trata de una noción cerrada, sino en evolución y cambio constante y, por ende, ambigua y proclive a admitir múltiples interpretaciones.

2.2. IA: UN POCO DE HISTORIA

Si bien la IA tiene la consideración de tecnología moderna, ya en el papiro de Edwin Smith, calificado como el primer tratado quirúrgico de la historia, que data del año 1600 a.C. (aunque se considera copia de un papiro del año 3000 a.C. aproximadamente), constaban premisas con la estructura "*if…then*" ("si…entonces"), una de las bases de los algoritmos actuales.

Por su parte, en los mitos griegos y en la literatura árabe se hallan recurrentes referencias a la creación de seres autómatas o artificiales. En concreto, en Alejandría, a partir del siglo III a.C. apareció una escuela de científicos que inventó complejas y novedosas máquinas, incluyendo autómatas que reproducían actividades humanas[19], siendo las más célebres las creadas por Ctesibio de Alejandría, autor de la primera máquina autocontrolada, un regulador del flujo de agua, y Herón de Alejandría, autor del trabajo "Autómata", que incluía máquinas con capacidad de actuar en función de su diseño y mecánica sin requerir supervisión humana (por ejemplo, la primera máquina de vapor, una bomba para apagar incendios o un teatro de marionetas).

Ya en la época medieval, el alquimista suizo Paracelso anunció la creación de un "*homúnculo o ser artificial con nada más que su esperma, magnetismo y alquimia*". [20]

Especial relevancia tuvo, en los siglos XV-XVI, Leonardo Da Vinci, quien en su faceta de inventor creó para los reyes Luis XII y Francisco I de Francia el denominado "*león mecánico*", una figura capaz de moverse sin ayuda humana; y, asimismo, diseñó el

[19] National Geographic, 2017.
[20] Wilkins, 2019, pág. 284.

conocido como "*Automa cavaliere*", un caballero mecánico que "*bajo una armadura completa, ocultaba levas, poleas y engranajes para mover brazos y piernas*".[21]

Asimismo, en el siglo XVII, René Descartes, quien llegó a identificar el cuerpo humano con una máquina, al parecer, tras el precoz fallecimiento de su única hija, mandó construir una muñeca autómata, fiel reproducción de la misma, que fue llamada Francine[22]. Además, en su Discurso del Método, sentó las bases de lo que posteriormente resultaría ser el Test de Turing, manifestando (con subrayado propio): "*no parecerá de ninguna manera extraño a los que, sabiendo cuántos autómatas o máquinas semimovientes puede construir la industria humana, sin emplear sino poquísimas piezas, en comparación con la gran muchedumbre de huesos, músculos, nervios, arterias, venas y demás partes que hay en el cuerpo de un animal, consideren este cuerpo como una máquina que, por ser hecha de mano de Dios, está incomparablemente mejor ordenada y posee movimientos más admirables que ninguna otra de las que puedan inventar los hombres. Y aquí me extendí particularmente, haciendo ver que si hubiese máquinas tales que tuviesen los órganos y figura exterior de un mono o de cualquier otro animal, desprovisto de razón, no habría medio alguno que nos permitiera conocer que no son en todo de igual naturaleza que esos animales; mientras que si las hubiera que semejasen a nuestros cuerpos e imitasen nuestras acciones, cuanto fuere moralmente posible, siempre tendríamos dos medios muy ciertos para reconocer que no por eso son hombres verdaderos; y es el primero, que nunca podrían hacer uso de palabras ni otros signos, componiéndolos, como hacemos nosotros, para declarar nuestros pensamientos a los demás, pues si bien se puede concebir que una máquina esté de tal modo hecha, que profiera palabras, y hasta que las profiera a propósito de acciones corporales que causen alguna alteración en sus órganos, como verbi gratia* (…) *no se concibe que ordene en varios modos las palabras para contestar al sentido de todo lo que en su presencia se diga, como pueden hacerlo aun los más estúpidos de los hombres; y es el segundo que, aun cuando hicieran varias cosas tan bien y acaso mejor que nosotros, no dejarían de fallar en otras, por donde se descubriría que no obran por conocimiento, sino solo por la disposición de sus órganos* (…)".[23]

[21] Véase Bejerano, 2019.
[22] Véase Glez, 2018.
[23] Descartes, 1637, págs. 51-52.

Posteriormente, en el siglo XVIII, Europa quedó entusiasmada con un supuesto ser autómata llamado "El Turco" creado por el inventor eslovaco Wolfgang von Kempelen, capaz de jugar al ajedrez y ganar a las grandes figuras de la época, aunque posteriormente ello resultó ser una farsa, ya que se descubrió que dicho personaje era manejado por un humano.[24]

En el siglo XIX, en concreto en la década de 1840, la matemática inglesa Lady Ada Lovelace, teniendo en mente la Máquina Analítica creada por su colega Charles Babbage en 1834, se avanzó al futuro y predijo que las máquinas "*podrían componer piezas musicales y científicas de cualquier grado de complejidad o extensión*" y podrían "*expresar los grandes hechos de la naturaleza*", haciendo posible "*una época gloriosa para la historia de las ciencias*."[25]

En la historia más reciente debemos remontarnos al año 1943, cuando los neurocientíficos americanos Warren S. McCulloch y Walter Pitts crearon el primer modelo neuronal moderno, que ha servido de punto de partida para el desarrollo de muchos de los modelos neuronales actuales.

Poco después, en 1948 tuvo lugar en el California Institute of Technology (Caltech) de Pasadena (California, EEUU) una conferencia interdisciplinar, con título "*The Hixon Symposium on Cerebral Mechanisms in Behavior*", en que se analizó cómo el sistema nervioso controlaba el comportamiento humano y cómo el cerebro podía ser comparado con un ordenador.[26]

Posteriormente, en el año 1950, Alan M. Turing, considerado el padre de la informática moderna por su creación de la conocida "máquina de Turing" *(*"*mecanismo teórico para modelar cualquier operación de computación*"[27]), publicó en la revista *Mind* un artículo titulado "*Maquinaria computacional e Inteligencia*", que arrancaba diciendo: "*Me propongo examinar la cuestión: "¿Pueden pensar las máquinas?*", habiendo establecido

[24] BBC News, 2018.
[25] Boden, 2017, págs. 16-17.
[26] Nilsson, 2009.
[27] López de Mántaras & Meseguer, 2017, pág. 18.

los cimientos de la denominada IA actual. Y es que el mencionado matemático no solo tuvo como objetivo la creación de una computadora que simulara la inteligencia humana, sino que además previó el rol tan fundamental que jugaría el aprendizaje automático en el desarrollo de la IA, asegurando que, en vez de intentar crear una máquina que emulara directamente la inteligencia madura de un adulto, quizás lo más adecuado sería intentar imitar la mente de un niño, sometiendo a la máquina a un
posterior proceso de aprendizaje.[28] Asimismo, Alan M. Turing fue pionero en el diseño de un programa informático para jugar al ajedrez.

No obstante, una de las grandes aportaciones de Alan M. Turing al mundo tecnológico, sin duda, fue la creación del denominado "test de Turing", calificado como la prueba de fuego de la IA. Y es que en virtud de dicho test, uno o varios humanos formulan preguntas a través de una pantalla de ordenador y, únicamente en caso de que estos no puedan distinguir si quien les contesta es un ser humano o una máquina, puede considerarse que esta es inteligente y que, por ende, ha pasado el test con éxito. Desde que dicho matemático diseñó tal test ningún programa informático había conseguido superarlo, si bien en 2014, un *chatbot* -robot programado para entablar y mantener conversaciones *online*- llamado Eugene Goostman, consiguió hacer creer al 33% de los jueces humanos que era "genuinamente humano".[29]

En 1955, como consecuencia del interés que empezaba a despertar en el ámbito científico la posibilidad de trasladar a una máquina el funcionamiento de la mente humana, se celebró en Los Ángeles (California, EEUU), en el contexto de la "*Western Joint Computer Conference*", una sesión sobre aprendizaje maquinal ("*Session on Learning Machines*"), en que se presentaron cuatro trabajos: tres sobre reconocimiento de patrones (por Wesley Clark, Belmont Farley, Gerald P.Dinneen y Oliver Selfridge, todos ellos del MIT Lincoln Laboratory, Massachusets, EEUU) y uno sobre máquinas capaces de jugar al ajedrez (por Allen Newell, un investigador de la Rand Corporation de Santa Monica, California, EEUU).[30]

[28] López de Mántaras & Meseguer, 2017, pág. 17.
[29] Fresneda, 2014.
[30] López de Mántaras & Meseguer, 2017, pág. 19.

No obstante, la expresión IA no se acuñó por primera vez hasta que en el mismo año John McCarthy[31], un informático estadounidense calificado como "padre de la IA", propuso la organización de una conferencia de verano en 1956 en la Universidad de Dartmouth (Nuevo Hampshire, EEUU), junto con Marvin Minsky, de la Universidad de Harvard (Massachusets, EEUU), Nathaniel Rochester, de IBM Corporation, y Claude Elwood Shannon, de Bell Telephone Laboratories, sobre dicho pionero concepto. Y es que fue precisamente en dicha conferencia, financiada por la Fundación Rockefeller, donde no solo se puso nombre a tal fenómeno tecnológico, sino que se establecieron las bases sobre las que este debía desarrollarse: "*El estudio debe basarse en la conjetura de que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede describirse con tanta precisión que existe la posibilidad de crear una máquina para simularlo*."[32] En dicha conferencia, además de discutir acerca del novedoso asunto planteado, Allen Newell y Herbert Simon presentaron un programa denominado "*Logic Theorist*" (LT) -conocido como "Máquina de la Teoría Lógica"-, que puso de manifiesto la vital importancia del procesamiento de símbolos y el uso de heurísticas en la resolución de problemas.

Después de la conferencia de Dartmouth aumentó significativamente el interés por la entonces novedosa tecnología recién denominada IA, y así, en noviembre de 1958 se celebró en el National Physical Laboratory de Teddington (Middlesex, Inglaterra) el primer simposio internacional sobre IA, denominado "*Mechanisation of Thought Processes*", donde John McCarthy presentó un innovador trabajo titulado "*Programs with common sense*".[33]

Poco tiempo después, en junio del 1959, se celebró en la sede de la UNESCO de París la "*First International Conference on Information Processing*" (IFIP), donde se presentaron diversos programas de resolución de problemas. Así, por un lado, Herb Gelernter introdujo un sistema con capacidad para explicar teoremas geométricos que ya empleaba heurísticas hoy muy utilizadas por los sistemas de IA; y, por otro lado, Herbert Simon, J.C. Shaw, y Allen Newell presentaron el conocido "solucionador general de problemas*"* ("*General Problem Solver*", más conocido como GPS), que trató de demostrar que las máquinas no solo eran capaces de solucionar problemas muy concretos, sino que podían enfrentarse a

[31] Premio Turing en 1971.
[32] McCarthy, Minsky, Rochester & Shannon, 1955.
[33] Véase McCarthy, 1958.

cualquier problemática general imitando la conducta de los humanos al tratar de solucionar problemas, principalmente lógicos.[34]

No obstante, pronto se evidenció que el enfoque de los creadores del GPS resultaba de difícil aplicación práctica, habida cuenta de que las problemáticas existentes aumentaban de forma significativa y la lista de posibles combinaciones de búsqueda lo hacía en consonancia. Así, los investigadores se pusieron manos a la obra para hallar nuevas formas de gestionar y tratar datos artificialmente representando el conocimiento humano.

Procede advertir que en un principio los expertos se limitaron a aplicar la IA al mundo de la lógica y de los juegos[35], especialmente el ajedrez, habiendo este desempeñado un rol fundamental en el desarrollo de tal tecnología.

Entre otros, es merecedora de atención la tarea de Arthur Samuel[36], quien en 1959 creó un programa de juego de damas que podía aprender y mejorar sus capacidades a base de practicar, pudiendo incluso jugar contra sí mismo a modo de entrenamiento.

Especialmente relevante fue, asimismo, la creación de Deep Blue por IBM, empresa que contrató a un grupo de investigadores de la Universidad Carnegie Mellon (Pensilvania, EEUU) para que desarrollara un programa capaz de jugar al ajedrez, habiendo llegado a vencer en un torneo de seis partidas (aunque solo en la última) celebrado en 1997 al campeón mundial Gary Kaspárov.[37]

Por su parte, el 14 de enero del 2011, tras siete años de desarrollo, IBM presentó en el concurso de televisión estadounidense Jeopardy! a Watson, una supercomputadora con miles de millones de bytes de memoria y 10 servidores de dicha compañía tras su esqueleto metáico -con tamaño equivalente a varios refrigeradores-, que ganó a Ken Jennings y Brad Rutter, los dos mejores concursantes de la historia del programa.[38] Tras ello, en 2014, IBM anunció que iba a comercializar la tecnología empleada con Watson promoviendo su

[34] Véase Newell & Simon, 1961, págs. 109-124.
[35] Véase Samuel, 1959, págs. 210-229.
[36] Samuel, 1959, págs. 210-229.
[37] Tran, 2021.
[38] Véase Markoff, 2011.

aplicación en distintos ámbitos (entre otros, gubernamentales y comerciales)[39], lo que ha generado un gran volumen de negocio a la compañía, que sigue mejorando y ofreciendo dicha tecnología entre su catálogo de productos.[40]

Finalmente, AlphaGo, desarrollado por Google DeepMind, fue el primer programa informático que el 9 de marzo del 2016 derrotó en Seúl (Corea del Sur) al campeón mundial de Go, Lee Sedol. No obstante, su supremacía fue breve, ya que la nueva versión, AlphaGo Zero, lanzada poco más de un año más tarde, venció al programa original por cien puntos a cero. La gran diferencia entre estos dos programas es que, en el primero, el sistema se entrenó a partir de datos relativos a miles de partidas jugadas por expertos humanos y, en cambio, en el segundo, el programa únicamente se ejercitó jugando repetidamente contra sí mismo.[41]

Con independencia de lo expuesto, de mediados de los años 50 a mediados de los años 80 hubo mucha actividad relacionada con la creación de sistemas capaces de aprender a reconocer patrones mediante redes neuronales artificiales, resultando hoy la aplicación más conocida la del *Machine Learning* o aprendizaje automático, y en especial, la del *Deep Learning* (una clase de Machine Learning, como se expondrá más adelante) o aprendizaje profundo.

Sin embargo, como consecuencia del desengaño producido por las limitaciones de la IA simbólica (que se definirá más adelante), que se había desarrollado con entusiasmo hasta el momento, pero que resultó ser demasiado rígida y poco dinámica, en 1970 tuvo lugar el denominado primer "invierno de IA", una época de parón en el ámbito de la investigación de tal tecnología.

No obstante, en los años 80 empezaron a proliferar proyectos de industrialización y comercialización de sistemas de IA, debiendo hacer especial mención al ambicioso "Proyecto de Sistemas Informáticos de Quinta Generación" (FGCS), impulsado en 1979 por el Ministerio de Industria y Comercio Internacional de Japón, que pretendía desarrollar

[39] IBM, 2018.
[40] Véase IBM, s.f.
[41] Knight, 2017.

sistemas de *hardware* y *software*, incluyendo ámbitos de *software* inteligente y, a pesar de que fracasó, sirvió para estimular el interés sobre la IA en el resto del mundo.

En tal línea de actuación se desarrolló en Europa y EEUU una tendencia relacionada con la creación de sistemas basados en conocimiento inteligente (IKBS) -también denominados sistemas expertos-, entre los que destacó el proyecto Alvey[42], en Reino Unido, y la Microelectronics and Computer Techonology Corporation (MCC), formada por un consorcio de compañías estadounidenses, para financiar proyectos de IA a gran escala.[43] Por su parte, la Agencia de Proyectos de Investigación Avanzados del Departamento de Defensa de los Estados Unidos (DARPA) asistió a la IA con el patrocinio de centros de investigación tales como el MIT Artificial Intelligence Laboratory, la Universidad Carnegie Mellon y Universidad de Stanford (y el Stanford Research Institute -SRI-)[44].

A partir de entonces, la investigación de la IA se centró en los ya mencionados IKBS -sistemas expertos-, que se distinguían del objetivo y la base de conocimiento universal del sistema GPS basándose en dominios de conocimiento específico para la resolución de problemas[45].

No obstante, tales sistemas pronto dejaron patentes sus limitaciones y fragilidades, y en tal contexto empezaron a cobrar fuerza los sistemas basados en redes neuronales.

Tales sistemas ya se habían ido desarrollando desde los años 40, primero en 1943 por Warren McCulloch y Walter Pitts, quienes expusieron una teoría sobre el modo de trabajar de las neuronas y diseñaron una red neuronal simple a partir de circuitos eléctricos; y, posteriormente, en 1949 por Donald Hebb, quien explicó por primera vez los procesos del aprendizaje desde un punto de vista psicológico, entendiendo que el aprendizaje tenía lugar cuando se activaban determinados cambios en las neuronas, habiendo lanzado así la denominada Teoría de Hebb.

[42] Véase Center for Computing History, s.f.
[43] Véase Teigens, Skalfist, & Mikelsten, 2019.
[44] Entre otros, un equipo de investigadores de SRI creó Shakey el robot, uno de los primeros vehículos autónomos.
[45] Véase Darlington, 2017.

Asimismo, numerosos científicos llevaron a cabo trabajos en tal sentido durante los años 50, entre otros, Marvin Minsky, quien en 1954 defendió su tesis doctoral en la Universidad de Princeton bajo el título "*Theory of Neural-Analog Reinforcement Systems & Its Applications to the Brain-Model Problem*"; Bermont Farley y Wesley A. Clark, del MIT, quienes en 1954 consiguieron crear las primeras simulaciones por computadora de pequeñas redes neuronales[46]; Nathaniel Rochester y John Holland[47], que hicieron lo propio en 1956; y W.K. Taylor de la University College de Londres (Inglaterra) o J.T. Allanson, de la Birmingham University (Inglaterra).

Especialmente relevante fue la tarea de Frank Rosenblatt, del Laboratorio Aeronáutico de Cornell (Nueva York, EEUU), quien en 1957 empezó a investigar redes neuronales artificiales a las que llamó "perceptrones", habiendo sido su trabajo alabado incluso en los medios de comunicación del momento por el impacto tecnológico que supuso [48], subyaciendo todavía a día de hoy en la base de novedosos proyectos[49], principalmente de identificación de patrones.

En 1960 Bernard Widroff y Marcian Hoff crearon el modelo Adaline ("*ADAptative LINear Elements*"), primera red neuronal aplicada a un problema real, que consistió en la creación de filtros para suprimir ecos en las líneas de teléfono, que fue empleada comercialmente durante varias décadas.

Unos años después, no obstante, en 1969 Marvin Minsky y Seymour Papert dieron un duro golpe a los sistemas de redes neuronales artificiales, ya que destaparon la debilidad del Perceptron de Rosenblatt, demostrando que no tenía capacidad para resolver problemas relativamente fáciles.

No obstante, a principios de los años 80 tuvo lugar un nuevo hito en el ámbito de la IA que hizo resurgir de forma exponencial el interés científico y económico por dicha tecnología: el desarrollo de los sistemas de aprendizaje automático o *Machine Learning*.

[46] Véase Farley & Clark, 1954, págs. 76-84.
[47] Rochester, Holland, Habit & Duda, 1956, págs. 80-93.
[48] The New York Times, 1958.
[49] Véase Emerging Technology from the arXiv, 2018.

En 1985, no obstante, John Hopfield contribuyó de nuevo al auge de las redes neuronales con la publicación de su libro: "*Computación neuronal de decisiones en problemas de optimización*.", lo cual empezó a abrir camino para la creación de numerosos trabajos y proyectos basados en redes neuronales artificiales, que han ido sucediéndose a lo largo de los años 90 y hasta la actualidad. Y es que, tal y como manifiesta José Dorronsoro Ibero, catedrático de Ciencias de la Computación e Inteligencia Artificial en la Universidad Autónoma de Madrid: "*Ahí resurgen las redes neuronales, todavía muy superficiales y limitadas por la disponibilidad de datos y los recursos técnicos, pero que permiten vivir la primera primavera del enfoque neuronal*".[50]

Si bien el concepto de *Deep Learning* o aprendizaje maquinal profundo ya fue extendido por los expertos informáticos de los años 80, en 1986 un equipo de investigadores, entre los que se hallaba Geoffrey Hinton, logró crear un método de entrenamiento basado en la "retropropagación" ("*backpropagation*") que abrió la puerta a las redes neuronales artificiales sin supervisión. Gracias al desarrollo de tal algoritmo de retropropagación se sentaron las bases del denominado aprendizaje profundo o *Deep Learning*, el más extendido en el ámbito de la IA actual, que ha superado con creces las limitaciones de la IA simbólica.

Hasta la actualidad, el *Machine Learning* y, en concreto, el *Deep Learning* ha sido la tecnología más empleada y desarrollada en los sistemas de IA, habiéndose afirmado desde IBM que hemos entrado en la "Era cognitiva", que "*marca la siguiente etapa de la aplicación de la ciencia para comprender la naturaleza y promover la prosperidad de la humanidad"*[51]. Y es que este tipo de sistema de IA es el responsable de que Google Maps nos guíe hasta nuestro destino buscando las rutas que más se adaptan a nuestras necesidades; que Microsoft y Google Translator traduzcan textos de un idioma a otro cada vez con más precisión; que Siri y Alexa reconozcan nuestra voz; y que Netflix nos sugiera películas y series acordes a nuestros gustos. En 2019, de hecho, el premio Turing (Turing Award), otorgado a la excelencia en el campo de la IA, fue otorgado a tres de los más

[50] Iglesias, 2019.
[51] Michael, 2016.

influyentes arquitectos del *Deep Learning*: Yann LeCun, de Facebook; Geoffrey Hinton, de Google; y Yoshua Bengio, de la Universidad de Monreal (Canadá).

De acuerdo con lo expuesto, resulta evidente que desde la Conferencia de Dartmouth se han hecho múltiples avances en el ámbito de la IA, pero lo cierto es que no ha sido hasta los últimos años cuando estos se han acelerado y han resultado más efectivos, habiendo hecho surgir la denominada "Cuarta Revolución Industrial". Ello, tal y como Microsoft puso de manifiesto, se debe principalmente a las siguientes circunstancias:

- ✓ "*la elevada disponibilidad de datos,*
- ✓ *la creciente potencia de la nube de los ordenadores,*
- ✓ *y la existencia de algoritmos más poderosos creados por investigadores de IA*."[52]

No obstante lo anterior, no todo ha sido bonanza, éxito y progreso en el ámbito de la IA, ya que además de los investigadores partidarios de tal tecnología, también han surgido detractores[53], habiendo incluso existido, tal y como se ha avanzado, algunos periodos en los que apenas han existido avances, los ya denominados "inviernos de IA", fundamentalmente en la década de los años 70, en que, especialmente crítico fue el denominado informe Lighthill[54], publicado en 1973 en el Reino Unido, que puso de manifiesto que ninguno de los descubrimientos realizados hasta el momento, en el campo de la IA, había provocado un impacto tan significativo como el prometido.

El futuro del fenómeno de la IA es, sin duda, incierto. Y es que mientras algunos consideran que habrá una gran revolución en el ámbito de tal tecnología durante este siglo, otros, los más prudentes, dentro de los que se incluyen Martin Rees[55], piensan que las habilidades humanas no serán superadas por la IA hasta dentro de varios siglos. Por el momento, la ya mencionada la Agencia de Proyectos de Investigación Avanzados del Departamento de Defensa de los Estados Unidos (DARPA) está trabajando en el proyecto Kairos, que tiene como objetivo interpretar la actualidad del mundo y predecir el futuro[56] .

[52] Smith & Shum, 2018, pág. 32.
[53] Dreyfus, 1965, pág. 3.244.
[54] Véase Lighthill, 1973.
[55] Rees, 2017.
[56] Véase Barbieri, 2019.

2.3. TIPOS DE IA

Como cuestión previa es importante reiterar que, debido a mi perfil lego en ciencia (y en concreto, en informática y matemáticas), las pinceladas que voy a ofrecer en este apartado sobre los distintos tipos y clases de IA existentes en la actualidad no tienen otro objetivo que el de establecer unas bases conceptuales mínimas para tratar de entender y contextualizar el grueso de esta investigación, de carácter eminentemente jurídico.

Así, si bien la doctrina científica ha realizado numerosas clasificaciones de tipos y subtipos de IA, entiendo que la siguiente recopilación abarca una muestra razonablemente completa de las mismas.

En primer lugar, atendiendo a su contenido y a su nivel de similitud con la inteligencia humana, debemos atender a la distinción entre IA fuerte o general/universal (*Strong AI*) e IA débil o específica (*Weak AI*), creada por el filósofo John Searle, profesor de la Universidad de Berkeley (California, EEUU), en un crítico artículo sobre IA publicado en 1980.[57]

Por un lado, la IA fuerte es aquella cuyo objetivo último es *"lograr que una máquina tenga una inteligencia de tipo general similar a la humana"*[58], definiéndola algunos autores como la "*clonación de la inteligencia humana*".[59]

A dicho tipo de IA es al que aspiraron los primeros científicos que se adentraron en el ámbito de tal tecnología (Alan M. Turing, John McCarthy y Marvin Minsky, entre otros) y con la que seguimos soñando hoy en día aunque, por lo que parece, todavía lo continuaremos haciendo durante unas cuantas décadas, habida cuenta de que su consecución es un reto, por el momento, imposible de alcanzar.

[57] Véase Searle, 1980.
[58] López de Mántaras & Meseguer, 2017.
[59] Rogel, Lacruz, Mozo & Diaz, 2018, pág. 36.

Y es que los pioneros de la IA pecaron de ser demasiado optimistas al respecto, lo cual ha sido duramente criticado por algunos autores a lo largo de los tiempos[60]. Con tal enfoque idealista y entusiasta, en 1955, John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude E. Shannon, afirmaron en el documento presentado para la celebración del proyecto de IA en la Universidad de Dartmouth durante el verano de 1956 ("*A proposal for the dartmouth summer research project on Artificial Intelligence*") que: "*las capacidades de velocidad y memoria de las computadoras actuales tal vez sean insuficientes para simular muchas de las funciones superiores del cerebro humano, pero el principal obstáculo no es la falta de capacidad de la máquina, sino nuestra incapacidad de escribir programas que aprovechen por completo lo que tenemos*…".

Dichos científicos tenían claro que la IA tenía capacidad para replicar de forma exacta y total la inteligencia humana, indicando, justamente, que si ello no ocurría no era por las eventuales limitaciones maquinales, sino por las humanas. No obstante, tales científicos no consiguieron que sus proyectos superaran con éxito ni siquiera el conocido test de Turing y, por ende, no alcanzaron los objetivos que de inicio se habían marcado, habiendo llegado a reconocer, incluso, en el año 2006, durante la celebración del 50 aniversario de la Conferencia de Dartmouth, que habían subestimado las dificultades de tal tecnología y habían pecado de idealistas y optimistas.[61]

No obstante, desde luego, cada vez estamos más cerca de alcanzar una IA general, sobre todo por la creciente potencia y capacidad de los sistemas informáticos. Sin embargo, ello no lo es todo, ya que se necesitan también nuevos métodos y estrategias que requieren "*heurísticas, planificación, simplificación matemática y representación del conocimiento*"[62]. Además, si bien la inteligencia humana es mucho más compleja de lo que podemos imaginar y abarca multitud de aspectos (cognitivo, emocional, sensorio-motriz y social), la mayoría de los actuales proyectos generalistas se centran única y estrictamente en la cognición. No obstante, al parecer, lo que complica el logro de una IA general no es

[60] Entre otros, el filósofo Hubert Dreyfus.
[61] Véase López de Mántaras, 2016.
[62] Boden, 2017.

el *hardware*, sino el *software*, y es que "*tenemos que encontrar los algoritmos correctos, pero nadie todavía se ha acercado ni siquiera a ello*".[63]

Así, por un lado, no es ni siquiera seguro que lleguemos a conquistar la tan ansiada IA general; y, por otro lado, no hay desde luego visos de que, en caso afirmativo, ello vaya a lograrse en los años próximos.

Algunos, no obstante, van más allá y no solo creen que llegaremos a conseguir una IA de carácter universal sino que vislumbran la llegada de la denominada IAS ("Inteligencia Artificial Sobrehumana"), que tendrá capacidades muy superiores a las nuestras e incluso llegará a desbancar a la especie humana. En relación con ello, surge la expresión "singularidad tecnológica", una hipótesis en cuya virtud se asegura que llegará un momento en que los propios sistemas de IA serán capaces de auto-mejorarse de forma recurrente, creando nuevos sistemas cada vez más inteligentes que los anteriores, que quedarán fuera de nuestro control, con los enormes riesgos que ello implica. En tal línea se llevan a cabo diversos proyectos, entre los que destacan el estudio elaborado por el Future of Humanity Institute de la Universidad de Oxford[64], y estudios en la Singularity University, impulsada entre otros por la NASA y por Google, lo que, de forma inevitable, provoca discusiones doctrinales relativas a complejas cuestiones, entre otras, la eventual personalidad jurídica de las máquinas, tal y como veremos más adelante.

Según aseguró en 1970 Brad Darrach, Marvin Minsky manifestó que "*Si los humanos tienen suerte, puede que* (las máquinas) *decidan conservarlos como animales de compañía. Si no tienen suerte, se les tratará como comida*"[65], a pesar de que posteriormente fue rebajando la intensidad de su discurso, habiendo incluso llegado a afirmar: *"¿Heredarán la Tierra los robots? Sí, pero serán nuestros hijos.*"[66]

Por su parte, el científico Stephen Hawking ya vaticinó en 2014 que "*la Inteligencia Artificial augura el fin de la raza humana*" [67], y por su parte, Elon Musk, creador de Paypal

[63] Chalmers, 2010.
[64] Véase Drexler, 2019.
[65] Darrach, 1970, pág. 66.
[66] Véase Minsky, 1994.
[67] Hawkins, 2014.

y consejero delegado de Tesla Motors, manifestó que "*la Inteligencia Artificial amenaza la existencia de nuestra civilización*" (Palazuelos, 2017).[68]

En concreto, es interesante hacer referencia al papel fundamental que la física cuántica podría jugar un en tal progreso, ya que tal y como afirma José Ignacio Latorre, Catedrático de Física Teórica en la Universidad de Barcelona y director del Centro de Ciencias de Benasque Pedro Pascual: "*la computación cuántica servirá para crear inteligencias artificiales todavía más potentes*."[69] y, asimismo, la tecnología 5G podría implicar un progreso en este sentido.

En relación con la tecnología 5G, que supone una auténtica revolución en el ámbito de las telecomunicaciones y aumenta considerablemente la capacidad de implementación y desarrollo de sistemas de IA a gran escala (las conexiones 5G multiplican por diez la velocidad de las 4G, y se prevé que en el futuro se alcancen velocidades doscientas cincuenta veces mayores[70]), existe una guerra abierta entre dos de los grandes gigantes tecnológicos mundiales: China y EEUU. Y es que, tal tecnología, que será sin duda empleada con éxito en el ámbito del vehículo autónomo, en el de las cirugías "teledirigidas" o en el de las *smart cities*, entre otros muchos, tiene la capacidad de otorgar enormes y suculentas facultades de control masivo de los dispositivos que la emplean, lo que, por un lado, despierta evidentes ansias de poder; y, por otro lado, suscita justificados temores ante terceros. Así, en un intento más de parar los pies a la administración china, en 2020 EEUU lanzó el veto a la compañía Huawei y manifestó claramente que el dominio asiático de las redes 5G supone una peligrosa amenaza para la seguridad y la economía estadounidense, habiendo incluso advertido a Europa de que, si permitía la entrada de Huawei, compañía considerada como "caballo de Troya" empleado por el gobierno y los servicios secretos chinos, tal y como manifestó el secretario de Defensa de EEUU, Mark Esper, [71] quedaría comprometido el futuro de la OTAN [72]. En concreto, además, EEUU amenazó a España

[68] Palazuelos, 2017.
[69] Véase Hernández, 2019.
[70] Muñoz, 2019.
[71] Agencias, 2020.
[72] Véase Benner, 2020.

con eliminar el intercambio de información de los servicios de seguridad e inteligencia en caso de que se permitiera la entrada de la tecnología 5G proveniente de Huawei.[73]

Hay expertos en IA, sin embargo, que niegan que tal progreso vaya a suceder, como por ejemplo Peter Bentley, quien asegura que la existencia de una "*Super IA general*" es pura ficción.[74] En tal sentido, ya en 1980 el filósofo estadounidense John Searle popularizó el experimento denominado "La habitación china", que pretendió poner de manifiesto que no resultaba posible que las máquinas llegasen a pensar o a procesar información como lo hacemos los humanos, tratando de demostrar que estas podían llegar a actuar sin comprender ni lo que estaban haciendo ni por qué lo estaban haciendo.[75]

En mi humilde opinión, habida cuenta de que carezco de conocimientos técnicos, siendo que la inteligencia propia de los seres humanos no solo se compone de "cuerpo" (capacidades intelectivas que residen en el cerebro) sino también de "alma" (concepto abstracto muy difícil de definir y, por ende, de imitar), va a resultar imposible la creación de un sistema de IA que sea equivalente de forma total y absoluta a la inteligencia humana o que incluso la supere. Y es que, tal y como reflexiona Miguel L.Lacruz Mantecón, Profesor de Derecho Civil de la Universidad de Zaragoza (con subrayado propio): "(…) *todos sabemos que la vida espiritual, la conciencia, es una característica o don de la humanidad que damos por supuesta en todo humano, y que creemos que también la tienen seres humanos que padecen graves enferemedades o están inconscientes (todos cuando dormimos, y a veces soñamos, y nuestras ideaciones oníricas son reales). Aunque no sepamos definirla e ignoremos cómo se produce, todos estamos de acuerdo en que se trata de una cualidad exclusivamente humana* (…)".[76]

No obstante, en virtud de todo ello, la Unión Europea decidió ya el 2017 establecer unos criterios y bases de protección de la especie, ante las numerosas voces que advertían de que la sustitución de los humanos por "robots" que superarían nuestra inteligencia estaba por llegar. Así, la Resolución del Parlamento Europeo, de 16 de febrero de 2017, con

[73] Benedito, 2020.
[74] Bentley, Brundage, Häggström & Metzinger, 2018.
[75] Plasencia, 2019.
[76] Rogel, Lacruz, Mozo & Diaz, 2018, pág. 76.

recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, estipuló en su Introducción:

"(…) *O. Considerando que la evolución en la robótica y en la inteligencia artificial puede y debe concebirse de modo que preserve la dignidad, la autonomía y la autodeterminación del individuo, especialmente en el ámbito de la atención y la compañía a las personas, y en el contexto de los dispositivos médicos que «reparen» o mejoren a los seres humanos; P. Considerando que existe la posibilidad de que a largo plazo la inteligencia artificial llegue a superar la capacidad intelectual humana*;",

y prosiguió estableciendo unos "*Principios generales relativos al desarrollo de la robótica y la inteligencia artificial para uso civil",* entre los que se hallaba:

"*3. Pone de relieve que el desarrollo de la tecnología robótica debe orientarse a complementar las capacidades humanas y no a sustituirlas; considera fundamental garantizar que, en el desarrollo de la robótica y los sistemas de inteligencia artificial, los seres humanos tengan en todo momento el control sobre las máquinas inteligentes; estima que debe prestarse especial atención al posible desarrollo de un vínculo emocional entre seres humanos y robots —especialmente en el caso de grupos vulnerables, como niños, personas mayores y personas con discapacidad—, y destaca los problemas que pueden plantear las graves consecuencias físicas y emocionales que este vínculo emocional podría causar a los seres humanos*."

Por otro lado, la IA débil, que es la única que conocemos hoy en día, es aquella que únicamente alcanza representar partes específicas y claramente definidas de las funciones inherentes a la inteligencia humana.

Las áreas de aplicación de este tipo de IA son altamente diversas, resultando las más destacadas la militar (por ejemplo, en "*misiones de reconocimiento de rutas terrestres, neutralización de artefactos explosivos improvisados y de munición que no haya explosionado*"[77]), la medicina (por ejemplo, con sistemas de diagnóstico), el *marketing*

[77] Instituto Español de Estudios Estratégicos, 2010, pág. 113.

(por ejemplo, con herramientas de reconocimiento de patrones para el envío de ofertas, con consiguiente fidelización de clientes), el transporte (por ejemplo, con los vehículos autónomos), las finanzas (por ejemplo, con sistemas de previsión y gestión de riesgos), las telecomunicaciones (por ejemplo, con administradores de congestión de redes de radio), el control de fronteras (por ejemplo, con sistemas de reconocimiento facial en los aeropuertos), y, desde luego, la investigación de delitos, lo cual será analizado de forma más profunda en el Capítulo Tercero.

Y es que la IA, hoy en día, está muy presente en nuestras vidas, de hecho, mucho más de lo que nos podemos llegar a imaginar.

En el ámbito del *marketing*, un ejemplo curioso e impactante del que se hizo eco en febrero del 2012 el periódico estadounidense The New York Times[78], a raíz de una investigación llevada a cabo por Charles Duhigg sobre la forma en que las empresas detectan y predicen nuestros comportamientos a través de los datos, es el de la empresa Target y su objetivo de analizar los hábitos de compra de sus clientes para detectar a mujeres embarazadas.

En concreto, en tal artículo periodístico se explica un caso ocurrido en Minneapolis (Minnesota, EEUU), en que la empresa Target empezó a mandar cupones de descuento de productos de bebé al domicilio de una adolescente que vivía con su padre, lo que provocó la indignación de este, habida cuenta de que no entendía qué sentido tenía aquello, siendo que su hija era menor de edad y no estaba en estado de buena esperanza. Ante tal situación, el mencionado progenitor se puso en contacto con la compañía Target para solicitar que cesara en los envíos de tales promociones, lo que motivó una disculpa por parte de tal empresa, que prometió averiguar lo que había sucedido. No obstante, poco después fue el propio padre de la adolescente quien remitió una disculpa a la empresa, al haber descubierto que si bien efectivamente su hija menor estaba embarazada, esta lo había mantenido oculto.[79]

Y es que ello no hace más que demostrar el increíble poder de la IA en la predicción y detección de comportamientos de los consumidores a partir de los datos: una empresa

[78] Véase Duhigg, 2012.
[79] Véase Lane, 2012.

puede llegar a conocer el embarazo de una adolescente antes que su propio padre, incluso aunque este conviva con ella.

En segundo lugar, por cómo se lleva a cabo el aprendizaje del sistema de IA, debemos distinguir entre sistemas de IA de aprendizaje supervisado, no supervisado y por refuerzo.

Por un lado, el aprendizaje supervisado se caracteriza por la presencia de un programador humano que introduce datos "etiquetados" (*labeled data*) y establece distintas variables de entrada (*input data*), en función de las cuales deberá asignarse una etiqueta de salida adecuada, mostrando al algoritmo las conclusiones a las que debe llegar, siendo por ende la salida del algoritmo, previamente conocida.

La forma de funcionar es similar a la de un maestro con un niño: se requiere que los posibles resultados del algoritmo sean conocidos y que los datos empleados para entrenar o enseñar al algoritmo estén previamente etiquetados. Así, una vez definida la función, con la introducción de nuevos datos no etiquetados, el algoritmo será capaz de asignarles la etiqueta correcta.

En el ámbito humano, un ejemplo sería el del profesor que quiere enseñar a sus alumnos cómo detectar la diferencia entre una camisa, una camiseta y unos pantalones, asumiendo que estos no tienen ningún conocimiento previo sobre dichas prendas.

Así, imaginemos que, con tal fin, el mencionado profesor se presenta un día en clase con varias prendas de ropa y dice a los alumnos que deben averiguar si cada una de esas piezas es una camisa, una camiseta o un pantalón. No les dice cómo deben ponerse de acuerdo, pero les dice que toda la clase tiene que dar un solo resultado para cada prenda, y a tal efecto pueden discutir, debatir, votar, repartirse las decisiones por partes, etc.

Ante tal escenario, con alta probabilidad, al principio los niños escogerían aleatoriamente y el profesor les daría el porcentaje de aciertos. Si tal porcentaje fuera alto, los alumnos sabrían que tendrían que cambiar pocos elementos decisorios, pero si fuera bajo, estos intuirían que deberían realizar varios cambios. Del mismo modo, si siempre que hicieran caso a Bernardo, por ejemplo, el porcentaje de éxito fuera bajo, aprenderían que el peso de

su opinión debería disminuir; mientras que si al hacer caso a Alicia, por ejemplo, la puntuación fuera siempre muy alta, entenderían que deberían hacerle más caso.

Y al final, a base de repetir el mencionado proceso muchas veces, los niños aprenderían a decidir a quién habría que escuchar más, tal vez establecerían un sistema de voto por mayoría entre los alumnos más fiables, quizás descubrirán que hay niños que clasifican mejor prendas oscuras quelas claras, etc, y, como clase, conseguirían un método que probablemente lograría clasificar con bastante precisión las camisas, las camisetas y los pantalones, sin que ninguno de los alumnos, no obstante, tuviera la visión completa.

Y es que, en general, el aprendizaje supervisado se basa en conocer de antemano el resultado esperado, compararlo con el resultado obtenido, y ajustar los parámetros del modelo con más o menos intensidad en función del grado de diferencia entre ambos.

En el ámbito del *Machine Learning*, un claro ejemplo de sistema de aprendizaje supervisado es el que detecta el denominado correo no deseado o "spam", que determina según los parámetros que el programador ha dispuesto previamente (remitente, asunto, si el el destinatario forma parte de una lista o no, etc), si el correo electrónico debe ir o no a dicha carpeta.

Por otro lado, el aprendizaje no supervisado se caracteriza por la ausencia de un programador que introduce datos etiquetados (*labeled data*) y parámetros de identificación, siendo el propio algoritmo el que detecta la repetición de patrones y lleva a cabo un auto entrenamiento y aprendizaje, sacando sus propias conclusiones sin indicación externa alguna. Así, no se requiere un programador que etiquete los datos previamente, determinando qué agrupaciones o patrones existen y cuál debe ser el resultado de tal interpretación, sino que el propio sistema lo hace por sí mismo.

Y es que la principal diferencia con los sistmas anteriores es que estos no disponen de datos etiquetados y, por ende, no se conoce previamente la salida deseada del modelo. No obstante, sí que es posible disponer de un gran número de *inputs* adicionales.

Así, por ejemplo, el sistema puede no saber que el objeto que analiza es un coche, pero igualmente puede pesarlo, medirlo, obtener el color, etc. O bien, en el caso de que un sistema, por ejemplo, quiera hacer segmentación de los clientes de un supermercado con tarjeta de fidelización, podría, a partir de los datos del historial de transacciones de esos clientes, un sistema de Machine Learning no supervisado, averiguar que hay 5 grupos de clientes diferenciados y que los parámetros clave que separan a los grupos entre sí son la frecuencia de las transacciones, el valor medio de la compra, y la hora en la que realizan la mayoría de sus compras.

Y es que en el ámbito del *Machine Learning*, un claro ejemplo de sistema de aprendizaje no supervisado es el que detecta los gustos de un determinado consumidor/cliente de una empresa a través de los datos aportados por sus compras, lo cual resulta muy útil para remitirle futuras ofertas y asegurar así su fidelización.

Los principales algoritmos de aprendizaje no supervisado son: algoritmos de agrupamiento (clustering), análisis de componentes principales (PCA), *Singular Value Decompotition* (SVD), y análisis de componentes independientes (ICA).

En el ámbito humano, siguiendo el ejemplo del caso anterior, el profesor diría a los alumnos que hay tres grupos de objetos, sin darles información alguna sobre cada grupo (aunque algunas técnicas sin supervisión parten de que se tome una decisión sobre el número de grupos, no obstante) y los alumnos, a través del análisis y combinación de patrones, podrían llegar a clasificar las prendas de ropa en el grupo correspondiente, pero sin saber que se trata de una camisa, una camiseta o un pantalón ya que en ningún momento se les habría dado información al respecto ni se les habrían comunicado porcentajes de éxito.

La forma de operar de este tipo de algoritmos suele estar basada en dos operaciones distintas, en ocasiones combinadas: agrupación (*clustering*) y reducción de dimensionalidad.

Por una parte, la técnica de agrupación o *clustering* tiene como misión la de clasificar elementos en distintos grupos en función de sus características. Así, se lleva a cabo un proceso de organización/agrupación de aquellos elementos que cuenten con características similares en distintos en grupos, que son denominados "*clusters*".

La clave en estos casos es el concepto de "distancia" entre elementos. Y es que, generalmente, lo que se hace es traducir cada elemento en una representación numérica, lo que a menudo va acompañado de una reducción de dimensionalidad (ya que si se trabaja con elementos que tienen miles de parámetros el proceso puede ser altamente engorroso, pero si son unas pocas decenas, mucho mejor). Y una vez se tiene cada elemento en formato numérico, puede calcularse lo parecidos o diferentes que dos elementos son entre sí.

Así, el objetivo de la clusterización es conseguir agrupar todos los elementos de tal modo que se minimice la distancia entre elementos del mismo "*cluster*" o grupo pero se maximice la distancia entre elementos de "*clusters*" distintos (o dicho de otra forma, que los elementos del mismo "*cluster*" sean lo más parecidos posible, y que los elementos de *clusters* distintos sean lo más diferentes posible.) Y el objetivo de la reducción de dimensionalidad es la rebaja del número de variables a tener en consideración en el análisis de los datos, tal y como ya se ha avanzado.

Y finalmente, el aprendizaje por refuerzo (en inglés, *Reinforcement Learning* -RL-)

Este tipo de IA se basa en la máxima "prueba-error", existiendo una interacción permanente con el sistema, mediante el envío a este de mensajes/notificaciones que le informan de si es correcto o incorrecto lo que va realizando, pudiendo incluso establecer una puntuación, a modo de evaluación. Así, tras la toma de muchas decisiones y sus correspondientes valoraciones/resultados, el sistema es capaz de extraer conclusiones y determinar qué caminos son los que con más probabilidad, llevan al éxito. Y, por ejemplo, en un sistema que escribe artículos para una revista, el refuerzo se basaría en si la revista los acepta o no.

Un ejemplo de sistemas que emplean este modelo de aprendizaje son los relacionados con el juego, como por ejemplo el ya anteriormente mencionado AlphaZero, de DeepMind.

Los principales algoritmos empleados en el aprendizaje por refuerzo son: Q-Learning, programación dinámica (*dynamic programming*), y SARSA (*State-action-reward-state-action*).

En tercer lugar, en función de la forma de representación del conocimiento y procesamiento de la información, puede distinguirse entre IA simbólica (enfoque clásico) e IA neuronal (también denominada subsimbólica o conexionista).

Por un lado, la IA simbólica, principalmente defendida y desarrollada durante los primeros años de historia de la IA, especialmente por Allen Newell y Herbert Simon, es aquella en la que el conocimiento se representa por símbolos que representan, en cierto modo, como una mente humana conceptualizaría el problema. Así, por ejemplo, si el sistema tiene por objeto estudiar un texto, este se representaría mediante sujetos, predicados, artículos, verbos, etc; y tiene como finalidad procesar rostros, estos se representarían mediante una nariz, un ojo, una boca, etc.

Dicha tipología de IA parte de que en el cerebro humano, el mundo real (situaciones, acciones, objetos, conductas o cualquier otra circunstancia) se representa mediante símbolos, calificados como discretos[80], que se combinan a través de reglas formales de tipo sintáctico para crear expresiones significativas.

De acuerdo con ello, a través de algoritmos, los sistemas de IA simbólica pueden aprender a comprender y llevar a cabo la manipulación de símbolos a partir de la información obtenida de los denominados sistemas expertos, en que la información y los símbolos se relacionan bajo la lógica condicional del "si… entonces" ("*if…then*").

Dichos sistemas expertos, de cuya calidad depende el éxito de la IA simbólica, funcionan con reglas lógicas tales como: 1. Todos los mares tienen agua. 2. El agua es navegable 3. X es un mar. 4. Así que X es navegable.

Si bien en un principio numerosos científicos se mostraron muy entusiastas y optimistas respecto de la IA simbólica, en la que se pusieron muchas expectativas, tal y como ya se ha expuesto anteriormente, poco a poco fueron haciéndose patentes sus limitaciones, habida cuenta de su rigidez y sus restringidas posibilidades de adquirir conocimiento de forma autónoma.

[80] Amoruso, Bruno & Dominino, 2007, pág. 338.

Un ejemplo de uso de sistemas de IA simbólica es el de la creación en 1996 del ya mencionado sistema de juego de ajedrez DeepBlue, por parte de IBM.

Por otro lado, la IA neuronal, principalmente desarrollada a partir de la mitad de la década de los 80, tras el primer invierno de IA, surgió como una revolucionaria motivación e inyección de energía para los científicos e investigadores de tal tecnología.
Esta tipología de IA es aquella en la que el conocimiento se representa por redes de neuronas artificiales que se conectan entre sí.

Esta tipología de IA parte de que en el cerebro humano, "*el conocimiento está segmentado en pequeñas unidades funcionales (neuronas artificiales) que se conectan con grupos más grandes (enfoque bottom-up o de abajo arriba)*"[81], creando redes neuronales, que se representan de modo artificial. Y es que, a diferencia de la IA simbólica, la IA neuronal no necesita "entender" lo que procesa, y los elementos son conjuntos de números que se procesan a lo largo de muchas capas en una red neuronal, por ejemplo, donde su significado se pierde. Así, los sistemas que emplean este tipo de IA no representan ningún concepto superior, son simplemente números que cuando se agrupan mediante alguna operación, arrojan un resultado que funciona.

La gran ventaja de la IA neuronal frente a la IA simbólica es su capacidad de auto-aprendizaje automático (*Machine Learning* y *Deep Learning*), resultando un sistema infinitamente más flexible y dinámico, con muchas menos limitaciones y más posibilidades que esta última.

Un ejemplo de uso de sistemas de IA neuronal es el de la creación en 2016 por Google Deep Mind del ya mencionado sistema de juego de Go, AlphaGo.

Los más optimistas, entre ellos Robert Hastrow, en relación con lo anterior, afirman que la similitud entre los circuitos de una computadora y el entramado neuronal humano, permitirá incluso que el contenido del cerebro humano sea transferido a las máquinas.[82]

[81] Ionos, 2020.
[82] Jastrow, 1985, pág. 176.

Mucho se ha escrito y debatido sobre las bonanzas y limitaciones/riesgos de ambos tipos de IA, habiendo existido cierta "guerra" doctrinal al respecto. No obstante, José Dorronsoro Ibero, catedrático de Ciencias de la Computación e Inteligencia Artificial en la Universidad Autónoma de Madrid, resumió así, bajo mi punto de vista, muy acertadamente, la relación histórica y lógica entre ambas tipologías de IA: "*Algunos defendían la visión simbólica, basada en el cálculo lógico, y otros el foco conexionista, que requería entender cómo funcionan físicamente las neuronas para replicarlas artificialmente. Y como en aquel momento la capacidad de cómputo era muy limitada, el enfoque conexionista estaba condenado al fracaso porque necesitaba demasiados recursos, por lo que todo el trabajo se centró en la línea simbólica*". [83] Y es que posteriormente, cuando tales recursos han llegado, la IA conexionista se ha hecho con el grueso de los sistemas de tal tecnología.

En cuarto lugar, atendiendo a su grado de profundidad, podemos distinguir entre el aprendizaje automático o *Machine Learning* y el aprendizaje profundo o *Deep Learning*.

Con carácter general, es importante poner de manifiesto que el aprendizaje profundo o *Deep Learning* es una tecnología que forma parte del aprendizaje automático o *Machine Learning*, que a su vez es un tipo de IA.

Así, por un lado, el aprendizaje automático o *Machine Learning* puede definirse como cualquier sistema que utiliza algoritmos para que una "máquina" o programa sea capaz de "aprender" a llevar a cabo ciertas acciones tales como clasificar, controlar, automatizar, etc.

Y, por otro lado, el aprendizaje profundo o *Deep Learning* es un tipo concreto de aprendizaje automático o *Machine Learning* que suele utilizar redes neuronales similares a las del cerebro humano para conseguir ese "aprendizaje", por lo que requiere mucha más potencia informática que el anterior. Y es que la razón por la que se califica de profundo o *deep* es porque los modelos de redes neuronales suelen tener muchas capas de "neuronas", es decir, son modelos muy "profundos".

[83] Iglesias, 2019.

Como ejemplo de aplicaciones de sistemas de aprendizaje automático o *Machine Learning* es interesante hacer referencia a aplicaciones y plataformas tales como Spotify, para escuchar música, o Netflix, HBO o Amazon Prime, para ver películas y series, que a través de algoritmos deciden qué canciones o qué largometrajes recomendar a los usuarios en función de las preferencias que han detectado.

Y, como ejemplo de aplicaciones de sistemas de aprendizaje profundo o *Deep Learning* resulta interesante hacer mención a los vehículos autónomos y los asistentes virtuales, a saber, entre otros Siri o Alexa, que pueden ejecutar comandos escritos o hablados y llevan a cabo tareas ciertamente complejas.

Finalmente, por su aspecto/forma exterior, podemos distinguir entre IA corpórea (robots) e IA incorpórea.

Por un lado, la IA corpórea es aquella que consta representada en soportes físicos, siendo los más conocidos y evidentes los que tienen forma similar a la humana, los conocidos como "robots androides o humanoides".

Las fabulaciones con robots vienen de lejos, siendo oportuno traer a colación, entre otros, el mito clásico de Pigmalión o la historia de Frankenstein, creada por Mary Shelley en 1818, con la influencia de "*las investigaciones de Galbani, Darwin y de Crosse sobre la posibilidad de insuflar vida a cuerpos inertes mediante el uso de la electricidad*".[84]

Existen multitud de robots, la mayoría humanoides, con distintos fines, si bien debe hacerse especial referencia a "Shakey" el primer robot móvil con habilidad para percibir y razonar sobre su ambiente, creado y desarrollado entre 1966 y 1972 por el SRI's Artificial Intelligence Center, y hoy expuesto en el Computer History Museum de EEUU[85]; así como a "Kodomoroid" y "Otonaroid", robots diseñadas en 2014 por el ingeniero japonés Hiroshi Ishiguro, y convertidas en las dos nuevas asistentes del museo Miraikan de Tokio[86]; y a "Sophia", una de las robots androides más modernas y completas que existen en la

[84] Rogel, 2018, pág. 15.
[85] SRI, s.f..
[86] Véase The Japanese Times, 2014.

actualidad, que fue creada en 2016 por el estadounidense David Hanson en el seno de su propia compañía, Hanson Robotics, y que tiene habilidad para mantener una conversación, gesticular, y aprender de cada interacción que realiza con los humanos, tal y como demostró en la conferencia tecnológica Talent Land celebrada en Guadalajara (México) en 2018.[87] Como curiosidad, en 2017 a "Sophia" le fue concedida la ciudadanía de Arabia Saudí, habiéndose convertido en el primer robot del mundo con ciudadanía reconocida.[88]

[89] Shakey

[90] Sophia

Por otro lado, la IA incorpórea es aquella que no consta representada en soporte físico alguno, sino que permanece en el ámbito "virtual".

Un claro ejemplo de sistema de IA incorpórea es Siri, la asistente virtual que Apple empezó a introducir en sus dispositivos en 2011 y que actualmente se ha convertido en un miembro más de la familia de millones de usuarios de tal compañía, siendo que a través del procesamiento del lenguaje natural responde preguntas, busca información, proporciona recomendaciones, cumple órdenes y controla la agenda.[91]

En relación con lo expuesto, se entiende necesario aclarar en este punto que, a pesar de que las expresiones "IA" y "robótica" suelen ser empleadas de modo indistinto, lo cierto es que, si bien están íntimamente relacionadas y se complementan, no tienen un sentido idéntico.

Así, una vez ya sentadas las bases conceptuales de la IA, procede poner de manifiesto que la robótica se define por la RAE como aquella "*técnica que aplica la informática al diseño*

[87] Corona, 2018.
[88] Véase BBC, 2017.
[89] SRI, s.f..
[90] Sharkey, 2018.
[91] Véase Apple, s.f..

y empleo de aparatos que, en sustitución de personas, realizan operaciones o trabajos, por lo general en instalaciones industriales.", ocupándose, básicamente, de todo lo relativo a los robots.

No resulta sencillo hallar una definición universal del término "robot", acuñado por primera vez por el escritor checo Karel Capek en el siglo XX[92]. En tal sentido, y siguiendo con lo dispuesto por la RAE, un robot se define como una "*máquina o ingenio electrónico programable que es capaz de manipular objetos y realizar diversas operaciones*." No obstante, una de las definiciones del término "robot" más completas que existen, bajo mi punto de vista, es la siguiente: "*Un robot es una máquina, provista de cierta complejidad tanto en sus componentes como en su diseño o en su comportamiento, y que manipula información acerca de su entorno para interactuar con él*". [93]

Interesante, asimismo, es la definición de robot (y de su relación con la IA) que se da en el "*Report of COMEST on Robotic Ethics*"[94], publicado el 14 de septiembre del 2017 por el World Commission on Ethics of Scientific Knowdelge and Technology (COMEST), perteneciente a la UNESCO, en el marco de la Organización de las Naciones Unidas, que dispone (con subrayado propio):

"*Los robots contemporáneos pueden caracterizarse por cuatro características centrales:*

- ✓ *<u>movilidad</u>, que es importante para funcionar en entornos humanos como hospitales y oficinas;*
- ✓ *<u>interactividad</u>, posible gracias a sensores, que recopilan información relevante del entorno y permiten que un robot actúe sobre el mismo;*
- ✓ *<u>comunicación</u>, hecha posible por interfaces de computadora o sistemas de reconocimiento y síntesis de voz; y*
- ✓ *<u>autonomía</u>, en el sentido de capacidad de "pensar" por sí mismos y tomar sus propias decisiones para actuar sobre el enotrno, sin control externo directo.*

[92] Instituto Español de Estudios Estratégicos, pág. 27.
[93] Barrio, 2018, pág. 38.
[94] Véase COMEST, 2017.

La robótica contemporánea generalmente incluye formas de IA: replicar la cognición y la inteligencia humana con sistemas informáticos, lo que resulta en máquinas capaces de hacer cosas que requieren una forma específica de inteligencia, como la capacidad de percibir y representar cambios en su entorno y planificar su funcionamiento en consecuencia. La IA es crucial para la autonomía del robot porque le permite realizar tareas complejas en entornos cambiantes y no estructurados, como conducir un automóvil y adaptarse a las condiciones en la carretera sin ser teleoperado o controlado por un ser humano.

Los robots realizan sus tareas a través de algoritmos: reglas o instrucciones para la solución de un problema. Se pueden distinguir dos tipos de algoritmos: algoritmos deterministas, que controlan el comportamiento predictivo de los robots deterministas; e IA o algoritmos estocásticos, con habilidades de aprendizaje que forman el "corazón" de los robots cognitivos. El comportamiento de un robot determinista, incluso si el robot es altamente complejo y autónomo (requiere poca o ninguna supervisión humana), está básicamente preprogramado y esencialmente determinado. Sin embargo, los robots cognitivos basados en IA aprenderán de experiencias pasadas y calibrarán sus algoritmos, por lo que su comportamiento no será perfectamente predecible y probablemente se convertirá en un problema digno de una seria atención y reflexión ética."

Así, acuerdo con lo expuesto, mientras la robótica es una técnica que emplea, para la consecución de un mismo objetivo (diseño, creación y desarrollo de robots), varias disciplinas convergentes, pero distintas entre sí (entre otras, informática, la electrónica, la mecánica, las telecomunicaciones, etc), la IA no es más que una de esas disciplinas o tecnologías empleadas a tal fin, sirviendo en concreto para dotar de un plus de sofisticación y autonomía a los robots, permitiendo que estos traten y procesen información de forma cada vez más autónoma y racional, dotándoles de una mayor independencia, singularidad y complejidad. Y es que tal y como dispone el periodista y escritor español Andrés Ortega: "*El robot, según lo ven algunos, es meramente el contenedor de la IA, mientras que esta es el software dentro del contenedor, que puede tomar decisiones. El robot no es en sí inteligencia artificial, pero tendrá, y tiene ya en muchos casos*." [95]

[95] Ortega, 2015, pág. 15.

Finalmente, a modo de curiosidad, para poner en valor la contribución a la robótica de los países de la UE, procede poner de manifiesto que, según se afirma en la Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de IA y robótica (Considerando E), alrededor de una cuarta parte de todos los robots industriales y la mitad de todos los robots de servicios profesionales existentes en el mundo están producidos por empresas europeas.

2.4. IA: POSIBLES BENEFICIOS Y POTENCIALES RIESGOS

2.4.1. LA ÉTICA Y LA MORAL

La capacidad humana para tomar decisiones de forma autónoma implica, sin lugar a dudas, la necesidad de recurrir a referentes éticos para resolver las problemáticas que continuamente se plantean. De acuerdo con ello, y partiendo de la base de que la IA tiene por objeto conseguir que las máquinas lleguen a llevar a cabo la misma clase de cosas que puede hacer la mente humana[96], la aplicación de esta tecnología nos plantea verdaderos retos en el ámbito de la ética y la moral.

¿Qué es la ética? Según la Real Academia de la Lengua Española (RAE), además de ser aquella parte de la filosofía que trata del bien y del fundamento de sus valores, la palabra "ética" hace referencia a aquello que se considera recto, conforme a la moral. Y ¿qué es la moral? Desde luego, una palabra con numerosas acepciones, si bien las más comunes son aquellas que hacen referencia, por un lado, a la doctrina del obrar humano que pretende regular el comportamiento individual y colectivo en relación con el bien y el mal y los deberes que implican y, por otro lado, a la actuación de una persona conforme con las normas que esta tiene del bien y del mal.

En cualquier caso, dejando de lado las definiciones conceptuales abstractas, procede poner de manifiesto que uno de los mayores retos espirituales con los que se ha topado la historia de la humanidad ha sido la concreción de qué valores y formas de actuar deben ser considerados éticos, conforme a la moral. Y es que tal cuestión, desde luego, no ha sido ni

[96] Boden, 2016, pág. 1.

será jamás resuelta de un modo uniforme y homogéneo, habida cuenta de la gran diversidad de culturas y civilizaciones que han convivido, conviven y convivirán en nuestro planeta.

En mi opinión, la moral hace referencia a aquella forma de actuar del ser humano consistente en garantizar su propio bienestar buscando siempre un equilibrio con el respeto del bienestar del resto de seres humanos, concebidos tanto a título individual como global, de modo colectivo. No obstante, no cabe duda de que existen tantas definiciones de moral como personas habitan el mundo, y ello fue justamente lo que, bajo mi punto de vista, creó en la humanidad la necesidad de alcanzar pactos sociales para convivir en paz, habida cuenta de lo inviable y salvaje que resulta para los humanos, como seres racionales que somos, vivir sin fijar unas bases comunes de convivencia que, compartidas por la mayoría, garanticen esos derechos individuales y colectivos a los que se ha hecho alusión, tal y como defendieron las teorías del contrato social desarrolladas por Hobbes, Locke y Rousseau.

En la actualidad, tales pactos sociales, existentes en mayor o menor medida y bajo diferentes formatos en prácticamente todos los lugares del mundo, tienen un contenido distinto en cada territorio. Sin embargo, existe un elemento común para todos los pueblos y naciones que, sin duda, marcó un hito en la historia de la humanidad: la Declaración Universal de Derechos Humanos proclamada por la Asamblea General de las Naciones Unidas en París, el 10 de diciembre de 1948.

No obstante, si bien ello fue un punto de partida extraordinario y se convirtió en una clara referencia que inspiró a múltiples países del mundo en la elaboración de sus normas constitucionales y sus leyes nacionales, lo cierto es que, al no ser un documento legalmente vinculante, todavía hoy existen grandes y significativas diferencias a la hora de considerar y tratar los derechos y libertades de las personas en los distintos territorios. Así, por ejemplo, mientras algunos pueblos consideran la pena de muerte una sanción conforme a la ética y la moral, otros entienden que constituye un ataque directo al derecho a la vida y a los valores que rigen en su sociedad; mientras algunas naciones entienden que la mutilación genital femenina resulta conforme con los principios morales, otras creen que ello va en contra del derecho a la libertad y la dignidad de la mujer; mientras algunas sociedades ven el aborto como una opción perfectamente conforme con la ética y la moral, otras imponen castigos penales a quienes lo practican; y, mientras algunas culturas perciben

la homosexualidad como una manifestación más del derecho a la libertad sexual de los ciudadanos, otras la persiguen con dureza.

Y es que es evidente que existen enormes diferencias culturales en el ámbito de las percepciones y/o los juicios morales y, por consiguiente, resulta interesante hacer referencia al binomio formado por el denominado universalismo moral (principios o estándares éticos universales) y el llamado relativismo cultural moral (normas éticas locales o culturales como fuente exclusiva de estándares éticos)[97]. En relación con ello, algunos autores defienden la premisa "*when in Rome do as Romans do*", es decir, "allí donde fueres, haz lo que vieres", si bien ello conduce inevitablemente al relativismo moral; otros entienden que procede actuar siempre conforme a los propios valores y normas morales, lo que en ocasiones es considerado como "imperialismo cultural"; otros defienden la idea de crear una moral universal [98]; y otros adoptan una posición intermedia[99] o buscan bases empíricas o filosóficas para la creación de una ética global[100].

Ello, tal y como se ha puesto de manifiesto, implica inevitables dilemas en el ámbito de las relaciones humanas, pero tal problemática aplicada al ámbito de la IA cobra especial relevancia. Y es que si bien hoy en día ya vivimos en un mundo globalizado en que las relaciones humanas han traspasado fronteras y están cada vez más internacionalizadas, lo cierto es que la creación y el desarrollo de las aplicaciones de IA se erigen como un punto de inflexión definitivo en el establecimiento de una red o sistema de relaciones globales, habida cuenta de los distintos estados de evolución en que se halla la investigación de dicha tecnología en los diferentes países del mundo, necesitándose unos a otros para avanzar en la buena dirección. Y ello se traduce, irremediablemente, en dificultades a la hora de determinar la posibilidad o no de utilización de ciertos sistemas o herramientas de IA en los distintos países en función de la base ética y moral que subyace en la configuración de estos. Así, ¿puede un sistema de reconocimiento facial creado en China, conforme a sus estándares éticos y legales, ser aplicado en un país europeo? ¿puede una empresa basada en EEUU utilizar un sistema operativo necesario para el desarrollo de un vehículo

[97] Melé & Sánchez-Runde, 2013, págs. 681-687.
[98] Por ejemplo, Michael Krausz.
[99] Por ejemplo, Thomas Donaldson o Jacob Gowans.
[100] Por ejemplo, Hans Küng.

autónomo creado por una empresa india, bajo su legislación? La polémica está servida, y desde luego no resulta obvia la solución, si bien en mi opinión, tal y como se verá posteriormente, deberían crearse agencias estatales (o, incluso interestatales, en el ámbito de la Unión Europea, por ejemplo, con delegaciones en los distintos territorios) que tuvieran por misión filtrar y analizar todos los sistemas de IA que pretendieran ser empleados dentro de sus fronteras y determinar si cumplen o no con unos mínimos estándares preestablecidos con arreglo a sus principios éticos y morales y, por supuesto, si son conformes con el ordenamiento jurídico imperante (tal y como ocurre ya con la Agencia Europea del Medicamento), no solo para garantizar la protección y el respeto a los derechos de sus ciudadanos sino también a los efectos de dejar bien sentadas las bases para la posterior imputación de eventuales responsabilidades.

En relación con ello, ya en la Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, se consideró necesaria la creación de un sistema global de registro de robots avanzados dentro del mercado interior de la UE, en los casos en que fuera pertinente y necesario para subcategorías específicas de robots, y se solicitó a la Comisión que estableciera criterios para la clasificación de los robots que tendrían que registrarse y que, justamente, analizase la conveniencia de que la gestión del sistema de registro y de las inscripciones se atribuyera a una agencia de la Unión para la robótica y la IA.

Por su parte, el 20 de febrero del 2020 el Consejo de Europa ya anunció que está estudiando la posibilidad de crear una Agencia Pública sobre IA, lo que, bajo mi punto de vista contribuiría, sin duda, a la irrupción, de forma definitiva y fiable, de la IA en las vidas de los ciudadanos europeos y, además, operaría como límite para varios países productores de sistemas de tal tecnología que no tienen unos estándares de protección de los Derechos Humanos tan altos como los nuestros, habida cuenta de que el mercado único europeo es el más grande del mundo y, por ende, la gran mayoría de empresas tecnológicas tienen un gran interés en poder operar en él.

Y es que, desde mi perspectiva, la creación de entidades de naturaleza pública, nunca privada, para el análisis y filtración de sistemas de IA con carácter previo a la concesión de autorización o licencia para su distribución y comercialización en un determinado

territorio, resulta fundamental. Y tal carácter público no solo resulta indispensable por la utilidad, la seguridad jurídica y la eficacia que implica la unificación de criterios, que también, y por la necesidad de minimización al máximo de intereses económicos individuales en beneficio de los colectivos, sino por la enorme sensibilidad y confidencialidad con que deben tratarse tales sistemas (habida cuenta de la gran cantidad de capital y recursos invertidos en la creación y el desarrollo de los mismos y las ingentes cantidades de datos, muchas veces personales, que albergan), que deben permanecer protegidos y blindados frente a los competidores, siendo que luego darán lugar a suculentas patentes. Y es que la única forma de avanzar en el binomio protección de derechos y libertades de los ciudadanos (libertad, seguridad, intimidad, etc) y protección de los derechos y libertades de las compañías (libertad de mercado y libre competencia, secreto de empresa, etc) es la de garantizar, con aval estatal o interestatal, de forma transparente y regulada, la absoluta confidencialidad, neutralidad y profesionalidad en el tratamiento de los sistemas de IA analizados, con previsión de un duro sistema de sanciones para los casos de inclumplimiento.

En tal sentido, tal y como veremos con detenimiento posteriormente, la Unión Europea ha presentado diversos documentos fijando las bases para la creación y el desarrollo de una IA competitiva y al mismo tiempo respetuosa con los valores y los derechos y libertades sobre los que se fundamenta tal comunidad. Así, en palabras de la Comisaria Europea para una Europa Lista para la Era Digital (2019-2024), Margrethe Vestager, "*queremos una economía competitiva y justa. Queremos mercados únicos, en las que empresas de todos los tamaños puedan competir en igualdad de condiciones, y donde el camino del garaje a la "startup" sea tan corto como sea posible. Pero esto también implica una economía en la que el poder del mercado que tienen algunas entidades no se puede utilizar para bloquear la competencia. Asimismo, va de una economía en la que los consumidores pueden estar seguros de que sus derechos se respetan y donde los beneficios pagan impuestos donde se producen.*"[101]

También existen iniciativas privadas, no obstante, que tienen por finalidad evitar que la IA resulte perjudicial para la humanidad, debiendo destacar especialmente Open AI, una

[101] Valdeolmillos, 2020.

organización sin ánimo de lucro creada en 2015 y financiada por varios emprendedores de renombre, incluidos Elon Musk, Peter Thiel y Sam Altman, destinada a estudiar las implicaciones éticas de la IA y a garantizar un desarrollo de la misma beneficioso para todos.[102]

Cynthia Breazeal, profesora del MIT Media Lab, manifiesta: "*Después de todo, la forma en la que experimentamos el mundo es a través de las comunicaciones y la colaboración, por lo que estamos interesados en máquinas que trabajen con nosotros, no podemos ignorar el enfoque humanista*". Y, por su parte, el científico y filósofo Collin Allen concluye: "*Tal y como podemos imaginar, desde la perspectiva humana, máquinas con crecientes grados de autonomía, podemos asimismo visualizar máquinas cuyos controles involucren grados crecientes de sensibilidad hacia las cosas que importan éticamente. No máquinas perfectas, sin duda, pero mejores.*" [103]

Respecto de todo lo anterior, entiendo que subyacen dos ideas o temores: por un lado, que los sistemas de IA sean programados por compañías basadas en países que no tienen unos niveles o estándares de protección de los derechos y las libertades de los ciudadanos tan garantistas como los propios; y, por otro lado, el de que los sistemas de IA lleguen a alcanzar un nivel de autonomía tal, que tomen sus propias decisiones sin contar con los humanos.

En relación con esto último, está claro que los sistemas de IA, por concepto, son entidades neutras, si bien no nacen solos, sino que son, en su origen, creados por seres humanos. No obstante, tal y como ya se ha ido avanzando, cada vez existen más sistemas de IA, sobre todo de *Deep Learning*, basados en el autoaprendizaje que, por ende, deben ser programados desde un inicio con unas directrices y unos modelos de actuación muy claros y precisos para asegurar que no se desvían de los objetivos fijados por sus creadores y no generan riesgos impredecibles y potencialmente nefastos.

En consonancia con ello surgen las polémicas preguntas: ¿las máquinas pueden pensar? ¿puede ser consciente una máquina? ¿puede sentir una máquina?. En mi opinión, en la

[102] Véase Knight, 2015.
[103] Nadella, 2016.

línea de las bases sentadas por el filósofo norteamericano John Searle, la respuesta a tales cuestiones, al menos por ahora, es negativa. Y es que los términos "pensar", "consciencia" y "sentimiento" tienen un contenido eminentemente subjetivo y ligado a la parte más íntima y sensible del ser humano, aquella que las máquinas todavía no han conseguido alcanzar. En tal sentido, el reputado científico estadounidense Jerry Kaplan, por un lado, manifiesta: "*Los programas informáticos, tomados por sí mismos, no cuadran realmente con nuestra intuición de sentido común sobre qué significa pensar. Solamente están llevando a cabo secuencias lógicas y deterministas de acciones, independientemente de lo complejas y cambiantes que sean sus configuraciones internas de un estado a otro*"[104] y, por otro lado, afirma: "*Es relativamente sencillo construir un robot que se retuerza de dolor, llore o simplemente diga que algo le duele cuando le pinchamos. Pero, como señala Peter Singer ¿dice eso algo sobre si siente dolor? Puesto que si podemos ir más allá de sus reacciones, a su estructura interna, la respuesta es que no. Reacciona de ese modo porque está diseñado para hacer eso, no porque sienta dolor*".[105] En relación con ello, John McCarthy investigó las posibilidades de que una máquina obtuviera el mayor grado de humanidad posible, el libre albedrío, habiéndose llegado a cuestionar en 1999*:"¿Podrá un ordenador decir algún día: puedo, pero no quiero?*".[106] Finalmente, John Markoff, en su libro "*Machines of Loving Grace*", dispone: "*El mejor modo de responder las duras y difíciles preguntas sobre el control en un mundo lleno de máquinas inteligentes es la comprensión de los valores de aquellos que justamente están creando tales sistemas*".[107]

Respecto de ello, es ya un clamor en el ámbito de la ingeniería la necesidad de colaboración estrecha con los operadores jurídicos en la fase de creación y programación de sistemas de IA, puesto que la mayoría se encuentran "solos ante el peligro", con miles de dudas sin resolver a la hora de introducir información en tales sistemas y establecer las bases para su funcionamiento. En tal sentido, los científicos tienen claro que no todo lo tecnológicamente viable es ética y jurídicamente posible y, por ende, necesitan unas guías claras sobre lo que pueden y no pueden hacer a la hora de crear y desarrollar programas de IA. Tal y como se ha avanzado y tal y como se pondrá de manifiesto más adelante, en la mayoría de paises

[104] Kaplan, 2017, pág. 79.
[105] Kaplan, 2017, pág. 90.
[106] Alandete, 2011.
[107] Markoff, 2016, pág. 27.

del mundo se hallan en el camino de regular tales cuestiones, si bien por el momento las herramientas jurídicas que existen son todavía muy abstractas y de difícil aplicación práctica, aunque como suele vulgarmente decirse, "*menos es nada*".

En relación con la cuestión ética y moral analizada, procede traer a colación las tres reglas de la robótica -disciplina, tal y como ya se ha expuesto, íntimamente relacionada con la IA-, calificadas como leyes fundamentales por Isaac Asimov, escritor de origen ruso y profesor de bioquímica en la Universidad de Boston (Massachussets, EEUU), que publicó en 1942 en el relato titulado "*Círculo vicioso*":

"*-un robot ni puede perjudicar a un ser humano ni puede con su inacción permitir que un humano sufra daño;*
-un robot debe obedecer las órdenes recibidas por un ser humano, excepto si tales órdenes entran en conflicto con la primera ley;
-un robot no debe proteger su propia existencia mientras tal protección no entre en conflicto con la primera o la segunda ley."[108]

Posteriormente, el mencionado autor, en su novela "Robots e Imperio", incluyó una cuarta ley fundamental, la denominada "ley cero", que debía imponerse a la primera:

"*-un robot no puede perjudicar a la humanidad, ni con su inacción permitir que la humanidad sufra daño*."[109]

En la actualidad, no obstante, ello no resulta del todo aplicable en el ámbito de la IA, puesto que tal y como ya se ha avanzado, está proliferando la creación de programas con capacidad de autoaprendizaje maquinal y, por ende, la obediencia al ser humano y su supervisión van quedando cada vez más difusas. Y es que, con razón, Stuart J. Russell, informático de origen inglés, profesor de la Universidad de Berkeley (California, EEUU) y de la Universidad de California, afirmó que el principal problema del control de la IA es cómo

[108] Barrio, 2018, pág. 55.
[109] *Idem.*

asegurarse de que los sistemas con un grado arbitrariamente alto de inteligencia permanezcan bajo estricto dominio humano[110].

En cualquier caso, la idea de que los robots -y, en general, los programas de IA- deben tener el objetivo de beneficiar a la humanidad y al mismo tiempo evitar perjudicarla, ha de s er, sin duda, el *alma mater* de tal tecnología y debe inspirar y servir como punto de partida a su regulación.

Y es que, bajo mi punto de vista, el peor riesgo que los sistemas de IA pueden entrañar es el de la deshumanización de la justicia.

En relación con ello, está claro que, al menos hoy en día, resulta imposible que un algoritmo, por muy potente que sea y por muy bien entrenado que esté, pueda tomar decisiones con la misma empatía y con la misma sensibilidad que un ser humano.

Así, desde mi propia experiencia como juez de instrucción, recuerdo de forma especial el caso de un senegalés que fue detenido por la Policía Nacional al salir de prisión, después de haber pasado diez años cumpliendo condena por la comisión de varios delitos contra la salud pública, y fue puesto a mi disposición para determinar si correspondía o no su ingreso en un Centro de Internamiento de Extranjeros (CIE) antes de ser expulsado del país, puesto que estaba aquí en situación irregular. Tengo muy presentes sus sentimientos de resignación y de desesperanza, puesto que él no esperaba tal desenlace a su salida de prisión. Recuerdo también a la perfección sus palabras: "*Durante los últimos siete años de mi vida, desde que logré desintoxicarme y deshabituarme de la droga, que es lo que me llevó a pasar por este infierno, solo he vivido con un sueño: poder abrazar a mi hija, con la que hablo cada semana, que tiene once años y vive en Alicante. Ahora mismo, pensar que puedo irme directo a un centro y luego ser expulsado a mi país, con prohibición de entrada en España durante los próximos cinco años, sin ver a mi hija, solo me provoca ideas suicidas y me quita todo el aliento para seguir viviendo*". A ello, el mencionado individuo añadió el terror que tenía de volver a caer en el mundo de las drogas si debía marcharse sin ver a su hija. Tales palabras, que sin duda desprendían una sinceridad

[110] Véase Rossell & Norvig, 2016.

cristalina, por cómo fueron pronunciadas, unidas al hecho de que había quedado acreditado que el mencionado individuo tenía una hija menor y que hablaba con ella semanalmente desde el centro penitenciario, me hicieron pensar mucho.

Respecto de ello, el artículo 62.1 de la Ley Orgánica 4/2000, de 11 de enero, de derechos y libertades de los extranjeros en España y su integración social, aplicable al caso, dispone que cuando una persona sea puesta a disposición del juez para decidir sobre su internamiento o no en un CIE, este (con subrayado propio): "*previa audiencia del interesado y del Ministerio Fiscal, resolverá mediante auto motivado, en el que, de acuerdo con el principio de proporcionalidad, tomará en consideración las circunstancias concurrentes y, en especial, el riesgo de incomparecencia por carecer de domicilio o de documentación identificativa, las actuaciones del extranjero tendentes a dificultar o evitar la expulsión, así como la existencia de condena o sanciones administrativas previas y de otros procesos penales o procedimientos administrativos sancionadores pendientes. Asimismo, en caso de enfermedad grave del extranjero, el juez valorará el riesgo del internamiento para la salud pública o la salud del propio extranjero*."

Así, es evidente que en estos casos la ley concede al juez un margen de actuación para que, guiado por los mencionados criterios, tome la decisión que considere más adecuada en cada caso concreto. Y en tal asunto, así fue.

Cierto es que el antedicho tipo no pudo probar de forma fehaciente que tenía familiares en Alicante que podían darle un hogar en cuanto llegara allí, principalmente porque estos residían en tal localidad y no resultaba viable tomarles declaración judicial en calidad de testigos en ese momento, si bien su Letrado reprodujo en la comparecencia una nota de audio en que se escuchaba una voz masculina que aseguraba ser el tío del individuo en cuestión y manifestaba que iría a recoger a su sobrino a la estación de autobús en cuanto llegara a Alicante, le proporcionaría un domicilio durante todo el tiempo que estuviera allí y se encargaría de acompañarlo a comunicar todos los nuevos datos de contacto a las autoridades.

Y, por todo ello, en aplicación del principio de proporcionalidad y la valoración de las circunstancias concurrentes, interpretadas y valoradas, sin duda, conforme a mi concepto de justicia (que, seguramente, no coincidiría con el de la "máquina", que habría estado entrenada para decidir conforme a criterios mucho más "estándar", basados en probabilidades y estadísticas), decidí no internar al mencionado joven en un CIE y le permití permanecer en libertad hasta que fuera expulsado a su país, puesto que entendí que otorgarle tal confianza iba a causar un bien mucho mayor, -tanto a él (que sin duda revivió cuando le notificaron la resolución), como a su familia (que iba a poder pasar tiempo con él antes de su marcha), como a la sociedad (que con alta probabilidad no iba a tener que sufrir su recaída en las drogas y la comisión de nuevos delitos)-, que el potencial riesgo o perjuicio que pudiera entrañar. Y tal ejercicio de reflexión es el que veo difícil que pueda llegar a hacer una máquina (al menos, en la actualidad).

De acuerdo con lo expuesto, los dilemas éticos en el uso de sistemas de IA son cada día más patentes, lo cual, inevitablemente, se traduce en el planteamiento de una problemática ligada a las eventuales responsabilidades que pueden surgir como consecuencia de las actuaciones de un sistema de IA contrarias a Derecho, a lo que se hará especial mención con posterioridad.

Pensemos, por ejemplo, en el ámbito militar, en el caso de los denominados "robots asesinos"[111], tales como el israelí "Harop", supuestamente empleado en la región de Nagorno-Karabaj (territorio disputado entre Azerbaiyán y Armenia) o el "SGR-A1", un robot desarrollado entre Samsung Techwin y la Universidad de Corea del Norte, que ya ha sido empleado para llevar a cabo tareas de vigilancia en la frontera entre Corea del Norte y del Sur, con capacidad de advertir la presencia de personas con la ayuda de sensores infrarrojos, con posibilidad de disparar sin asistencia humana[112]. La existencia de tales tecnologías, si bien puede implicar una alta eficacia estratégica en el ámbito militar, puesto que fundamentalmente trata de reducir al máximo el riesgo innecesario de los combatientes propios[113], puede conllevar riesgos éticos y morales elevadísimos contrarios a los derechos y libertades de los ciudadanos en aquellos casos de delegación absoluta de funciones a los

[111] Véase Travieso, 2015.
[112] Vilà, 2019.
[113] Instituto Español de Estudios Estratégicos, 2019, pág. 133.

algoritmos, habida cuenta de que las máquinas, cada vez con menos control humano, pueden llegar a cometer atrocidades tales como la de la ejecución de un menor, de uno de los propios aliados por error o del enemigo que, en realidad, pretendía rendirse.

En concreto, el ámbito del vehículo autónomo, la cuestión ética está siendo uno de los grandes desafíos de los investigadores y de las compañías productoras y comercializadoras. Y es que, el diseño de un vehículo sin conductor humano hace necesaria la creación de algoritmos preparados para afrontar una gran multitud de imprevistos que pueden ocurrir durante la conducción y tomar decisiones que, inevitablemente, implicarán compromisos éticos. En relación con ello, en 2018 un grupo de científicos europeos y estadounidenses realizaron un estudio denominado "*Moral machines*"[114], para determinar, mediante un videojuego en que los jugadores asumían el rol de vehículo autónomo, qué es lo que los humanos consideramos más o menos correcto ante los dilemas que se plantean al volante. Tal y como afirma Edmon Mad, el principal autor del estudio e investigador del Media Lab del MIT: "*Vimos que hay tres elementos que las personas tienden a aprobar más. Primero, entre salvar a un humano o un animal, el coche siempre debería atropellar a la mascota. La norma, además, primaría salvar al mayor número de personas. Así que si el conductor va solo y va a atropellar a dos peatones, que se estampe contra el muro. La tercera decisión más universal es que la mayoría cree que si un vehículo autónomo tiene que decidir entre chocar contra un niño o contra un anciano, el viejo debe morir para que el joven tenga la oportunidad de envejecer*."[115] No obstante, tal y como se deduce del mismo estudio, en el que participaron más de 2 millones de personas de distintas etnias y nacionalidades, las diferentes concepciones éticas varían según la cultura, lo que tiene una clara incidencia a la hora de tomar decisiones: así, por ejemplo, mientras los occidentales tenían claro en un altísimo porcentaje que ante la posibilidad de salvar a un menor o a una persona de la tercera edad, salvarían al primero, el porcentaje disminuía en el caso de las personas de etnia oriental, donde otorgan a la gente mayor un valor muy superior al que le atribuimos nosotros.

La clave, a la vista de lo expuesto, es que estos sistemas de IA puedan ser programados con ciertas limitaciones para garantizar el respeto a los valores, derechos y libertades de

[114] Véase Moral Machine, s.f..
[115] Criado, 2018.

cada sociedad, debiendo tener claro que el principal objetivo es el de obtener beneficios minimizando los riesgos que implica la intervención humana en determinadas circunstancias, lo que no puede, sin embargo, conseguirse "a cualquier precio". No obstante, Allen Newell, investigador americano que destacó en los campos de la IA y la cibernética, en relación con los procesos de toma de decisiones por sistemas programados, dispuso: "*cuando la decisión óptima es prácticamente imposible de determinar debido a que existe un número demasiado elevado de posibilidades a tener en cuenta, entonces se toma una decisión suficientemente satisfactoria aunque no sea óptima*". [116]

Procede poner de manifiesto, no obstante, que si bien las implicaciones y los desafíos éticos de la IA son sumamente relevantes en todos sus ámbitos de aplicación, existen sectores especialmente sensibles, por los derechos y libertades que hay en juego, tales como el militar, el sanitario y desde luego, el de la justicia, como analizaré posteriormente con más profundidad.

Así, si bien hoy en día todavía hay tareas y decisiones que por ahora solo la inteligencia humana puede asumir y tomar (en la línea de mi opinión de que las máquinas no pueden "pensar", "tener conciencia", o "sentir" como lo hacemos los humanos), se está avanzando a gran velocidad y está claro que tales limitaciones acabarán quedando reducidas a niveles prácticamente anecdóticos, aunque el cuándo todavía es una incógnita. La clave, en mi opinión, es que los creadores de sistemas de IA, desde un inicio, prevean "*los potenciales malos efectos que podrían venir de sus inventos*"[117], debiendo tender a la creación de las denominadas "*moral machines*".[118]

2.4.2. IA Y DERECHO

La Primera Revolución Industrial, iniciada en la segunda mitad del s.XVIII en Reino Unido y finalizada a mitad del s.XIX, estuvo marcada por la aparición de sistemas de producción mecánicos, con la invención de la primera máquina de vapor por James Watt, patentada en 1769, y la aparición de la locomotora de vapor, construida por primera vez de forma efectiva en 1814 por George Stephenson; la Segunda Revolución Industrial, que tuvo lugar

[116] López de Mántaras & Meseguer, 2017, pág. 7.
[117] Singer, 2009, pág. 426.
[118] Véase Wallach & Allen, 2010.

desde mediados del siglo XIX hasta principios del s.XX, fue motivada por la aparición de nuevas fuentes de energía tales como la electricidad (con la patente de la bombilla, inventada por Thomas Edison en 1880), el petróleo o el gas, innovadores sistemas de transporte, tales como el automóvil, y de comunicación, con inventos como el teléfono y la radio (con la primera transmisión en 1897), lo que motivó la aparición de nuevos sistemas de trabajo, como la producción en serie; y la Tercera Revolucion Industrial tuvo su origen en el planteamiento efectuado por Jeremy Rifkin[119], con la incorporación de microelectrónica y la tecnología de la información para automatizar la producción, con la aparición de los primeros ordenadores personales en 1962, el primer controlador programable (PLC) y la "*world wide web*" (Internet) en 1990.

Klaus Schaw, fundador del Foro Económico Mundial, en su libro "La cuarta Revolución Industrial" (Schaw, 2016)define la revolución tecnológica que estamos viviendo como la Cuarta Revolución Industrial, y dispone: "*Comenzó a principios de este siglo y tuvo como base la revolución digital. Está caracterizada por un Internet mucho más móvil y mundial, por sensores más pequeños y más potentes, y por inteligencia artificial y aprendizaje automático*".[120] Tal conclusión, sin duda, ha proliferado, habiendo sido el eje central de la reunión anual del Foro Económico Mundial celebrado en 2019 en Davos, Suiza, en que se analizaron los desafíos que presentan las nuevas tecnologías, entre las que se halla la IA, que está cambiando la forma en que interactuamos, trabajamos y vivimos.

Ante tal oleada de cambios, muchos son los que han alzado la voz para poner de manifiesto la necesidad de establecer guías y planes de actuación conjuntas con el objetivo de analizar y limitar los potenciales riesgos que entrañan las nuevas tecnologías y, a la vez, potenciar sus posibles beneficios, especialmente en el ámbito de la IA, habida cuenta del enorme impacto que esta puede tener en aspectos muy sensibles de la vida de los ciudadanos.

Así, entre otros, Nick Bostrom, Profesor de la Facultad de Filosofía de la Universidad de Oxford, manifestó en relación con la IA: "*Somos como niños jugando con una bomba*"[121], habiendo calificado la necesidad de regulación de tal tecnología como "*la tarea de nuestra*

[119] Rifkin, 2011.
[120] Schaw, 2016, pág. 37.
[121] Adams, 2016.

época". Por su parte, el sudafricano Elon Musk, cofundador de SpaceX, PayPal y Tesla Motors, entre otros, conocido por sus discursos sensibilizadores sobre los potenciales riesgos de la IA, ya en una ponencia pronunciada en el Centennial Symposium 2014 organizado por el Departamento de Astronáutica y Aeronáutica del MIT, calificó dicha tecnología como nuestra "*mayor amenaza existencial*" y, posteriormente, en 2017, en la charla ofrecida en la "*National Governors Association Summer Meeting*", recalcó la necesidad de ser proactivos y no reactivos en la regulación de la IA. Asimismo, el informático estadounidense Alan Kay advierte de que el futuro está en nuestras manos, manifestando que lo mejor es dejar de predecir cómo será y empezar a crearlo con unos principios claros desde ya.

Tal y como dispuso el comisario de Mercado Interior de la Comisión Europea Therry Breton (2019-2024): "*Reglamentar la IA es un poco como el Far West, empezamos en las tierras vírgenes y se hace un poco como se quiere, pero después hay que organizarse*"[122] y, por su parte, en la Introducción B de la Resolución del Parlamento Europeo de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho Civil sobre robótica se dispone: "(…) *resulta de vital importancia que el legislador pondere las consecuencias jurídicas y éticas, sin obstaculizar con ello la investigación*".

Y es que, ante la ya patente e incesante evolución de la IA, que se ha erigido como una tecnología con gran impacto estratégico y económico en nuestros tiempos, resulta más que evidente la necesidad de establecer una regulación (no excesiva, puesto que ello podría generar el efecto contrario al deseado) y delimitación sobre su creación, su uso y su comercialización.

Así, están proliferando iniciativas de carácter público y privado, a nivel mundial, que tienen por objetivo definir y establecer los límites y los principios sobre los que debe enmarcarse la regulación de la IA, habida cuenta de las amenazas que esta presenta para los seres humanos.

2.4.2.1. LA REGULACIÓN EN EL ÁMBITO PRIVADO

[122] Masdeu, 2020.

No cabe duda de que la regulación de la IA debe ser misión del sector público, puesto que habida cuenta del gran impacto social, económico y legal que presenta tal tecnología, las normas que imperen en dicho ámbito deben tener un carácter general y, sobre todo, de obligado cumplimiento. No obstante, por desgracia, tal y como ha mostrado la experiencia a lo largo de los años, el legislador es siempre el último en dar solución a las "nuevas" problemáticas que surgen en la sociedad, siendo siempre el sector privado el pionero en desarrollar iniciativas tendentes a poner límites y establecer los principios básicos que deben regir tales situaciones. Y en el caso de la IA, ello no ha sido distinto. Así, antes de que los riesgos y beneficios de la IA hayan cristalizado en legislaciones nacionales o transnacionales, numerosas empresas, asociaciones, fundaciones privadas, universidades y representantes de la sociedad civil, entre otros, han focalizado sus esfuerzos en sentar las bases de lo que se entiende que debe ser la respuesta pública a tal fenómeno emergente.

De acuerdo con ello, se reputa muy interesante, necesario e imprescindible para explicar y comprender mejor en qué punto estamos y, sobre todo, hacia dónde nos dirigimos en el ámbito de la regulación de la IA, hacer un análisis de las principales iniciativas y actuaciones llevadas a cabo por el sector privado en dicho campo.

Así, en el año 2016, un grupo de investigadores compuesto por Dario Amodei, Chris Olah y Dan Mané, de Google Brain, John Schulman, de OpenAI, Jacob Steinhardt, de la Universidad de Stanford (California, EEUU), y Paul Christiano, de la Universidad de Berkeley (California, EEUU), publicaron un informe titulado "*Concrete Problems on AI Safety*", en el que se abordaba, desde diversas perspectivas, el problema de los denominados accidentes de los sistemas de *Machine Learning*, entendiendo por tales aquellos comportamientos inintencionados y dañinos que pueden darse como consecuencia de la mala calidad del diseño de los sistemas de IA.[123]

Por su parte, la organización FAT ML (Fairness, Accountability, and Transparency in Machine Learning), que tiene por misión promover la creación y el desarrollo de sistemas de aprendizaje maquinal justos, explicables y transparentes, publicó un documento titulado

[123] Amodei & otros, 2016.

"*Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*"[124], que establecía las bases para ayudar a los creadores y comercializadores de tales sistemas de IA para su diseño e implementación de forma pública y sin opacidades, habida cuenta del gran impacto social que estos pueden tener.

Asimismo, en el seno de Future Tense -una colaboración entre Slate, New America y la Universidad Estatal de Arizona (EEUU) que explora las formas en que las tecnologías emergentes afectan a la sociedad, la política y la cultura-, se publicaron unas reglas sobre IA que básicamente ponían de manifiesto la necesidad de que esta fuera transparente, explicable, justa, segura y diseñada para ayudar a la humanidad, con maximización de sus eficiencias sin afectar a la dignidad de las personas.[125]

También en septiembre de dicho año, el "*Department of Computer Science*" de la Universidad Estatal de Florida (EEUU) publicó un informe en que estableció que el objetivo de las aplicaciones de IA debía ser, fundamentalmente, la creación de valor para la sociedad[126].

Y, finalmente, entre otras, Partnership on AI (PAI), una potente organización estadounidense que cuenta con reconocidos miembros de distintos ámbitos, tales como Accenture, Amazon, Access Now, AI Now Institute, de la Universidad de Nueva York (EEUU), Apple, Amnistía Internacional, la cadena BBC, Berkeley Center for Law & Technology, Berkman Klein Center, de la Universidad de Harvard (EEUU), Deep Mind, Google, Facebook, IBM, Human Rights Watch, Microsoft, MIT Media Lab, Samsung, Sony y Unicef, entre otros[127], y tiene por objeto desarrollar y compartir las mejores prácticas de IA, avanzar en el conocimiento y entendimiento de la misma por parte del público y erigirse como una plataforma de discusión y compromiso, hizo públicos ocho principios básicos que, bajo su punto de vista, deben regir su modo de trabajar e impregnar el ámbito de tal tecnología.[128]

[124] Véase FAT ML, s.f..
[125] Nadella, s.f..
[126] Véase Florida State University, 2016.
[127] Véase Partnership on AI, s.f..
[128] *Idem.*

En el año 2017, el Future of Life Institute (Cambdrige, Massachusetts, EEUU), centrado en "*catalizar y apoyar investigaciones e iniciativas para salvaguardar la vida y desarrollar visiones optimistas del futuro, incluidas formas positivas para que la humanidad dirija su propio curso, considerando las nuevas tecnologías y sus desafíos*"[129], con asesores del nivel de Elon Musk, Martin Rees, Stuart Russell o Nick Boström (y hasta su fallecimiento, Stephen Hawking), publicó los denominados "*Asilomar AI Principles*"[130], desarrollados en el marco de la "*Beneficial AI Conference*" (BAI) celebrada dicho año, que abogan por una IA segura, transparente, explicable, con valores humanos, priorizando el bien colectivo, con respeto a la privacidad y a la libertad, con control humano y fijación de un régimen de responsabilidades.

Por su parte, la Universidad de Montreal (Canadá) publicó "*The Montréal Declaration for a Responsible Development of AI*", estableciendo como principios básicos para el desarrollo y el uso de la IA, entre otros, el de bienestar, respeto de la autonomía, protección de la intimidad y la privacidad, la solidaridad, la participación democrática, la equidad, la inclusión, la prudencia, la responsabilidad y el desarrollo sostenible.[131]

Asimismo, la compañía estadounidense Intel publicó una serie de recomendaciones para el uso de la IA que incluían, entre otras, el fomento de la innovación y la apertura del desarrollo, la creación de nuevas oportunidades de empleo y la protección del bienestar, la liberación responsable de datos, siempre con respeto a la privacidad, y la responsabilidad por el diseño, con implementación ética.[132]

También Nicolas Economou, Consejero Delegado de la empresa H5, asesor de la "*AI Initiative of the Future Society*" en la Harvard Kennedy School (Massachusetts, EEUU), y defensor de la aplicación de métodos científicos al descubrimiento electrónico, proclamó una serie de principios sobre IA, basados en lo dispuesto en el "*Code of Practice for electronic discovery*" publicado (y revisado en el año 2020) por la International Organization for Standardizarion (ISO)[133], tales como la necesidad de su uso para mejorar el bienestar de la sociedad y la naturaleza, la transparencia y la explicabilidad, y la

[129] Future of Life, s.f..
[130] *Idem.*
[131] Véase Montreal Declaration, s.f..
[132] Véase Intel, s.f..
[133] Véase ISO, s.f..

conveniencia de que los sistemas de IA sean codificados a través de un diálogo inclusivo y reflexivo con la sociedad civil, entre otros[134].

Además, el Institute for Electrical and Electronics Engineers (IEEE) publicó, en el marco de la "*Global Iniciative on Ethics of Autonomous and Intelligent Systems*", tras la primera edición lanzada en 2016, la segunda versión del documento de principios generales "*Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*"[135], entre los que se hallaban el respeto y la promoción de los derechos humanos, la necesidad de aumentar el bienestar social, el empoderamiento de los individuos para acceder a sus datos y compartirlos de forma segura, manteniendo la capacidad de cada uno para controlar su identidad, efectividad, transparencia y explicabilidad, entre otros.

El Information Technology Industry Council (ITIC), autodefinido como "*la voz gobal del sector tecnológico*", hizo públicos sus "*AI Policy Principles*"[136], que ponen de relieve áreas específicas en que la industria, los gobiernos y otros operadores pueden colaborar y, asimismo, señalan oportunidades para llevar a cabo colaboraciones público-privadas. Entre otros, se hace referencia a la necesidad de un diseño y desarrollo responsable de la IA, con seguridad y control, así como a la existencia de datos robustos y representativos, la democratización del acceso a los sistemas, etc.

La reconocida sociedad académica japonesa Japanese Society for Artificial Intelligence (JSAI), por su parte, publicó "*The JASAI Ethic Guidelines*"[137] partiendo de la base de que la IA asumirá un papel importante en el futuro de la humanidad en una multitud de áreas, tales como la industria, la medicina, la educación, la cultura, la economía, la política, el gobierno, etc y que, sin embargo, es innegable que puede ser perjudicial para la sociedad o entrar en conflicto con los intereses públicos debido al abuso o a su mala utilización. Con base en ello, y en aras de garantizar que la investigación y el desarrollo de la IA resultan beneficiosos para los humanos, se establecieron una serie de principios a observar por los

[134] Economou, 2017.
[135] Véase IEEE, s.f..
[136] Véase ITIC, s.f..
[137] Véase The Japanese Society for Artificial Intelligence (JSAI), s.f..

miembros de la organización, que son, entre otros, la contribución a la paz, la seguridad, el bienestar y el interés público de la humanidad, el cumplimiento de las leyes y los reglamentos, el respeto por la privacidad ajena, la justicia, la seguridad, la integridad, la explicabilidad, la responsabilidad social y el comportamiento ético.

Asimismo, con ocasión de la celebración en Tokio (Japón) del taller de trabajo "*Beneficial AI*", organizado, entre otros, por el Next Generation Artificial Intelligence Research Center (AI Center) de la Universidad de Tokio (Japón), en asociación con el simposio sobre "*AI and Society*", organizado para explorar los desafíos de construir una comunidad global para garantizar un desarrollo seguro de la IA y beneficioso para todos, con un particular enfoque en Japón, los participantes expresaron su compromiso de trabajar hacia una IA beneficiosa para la sociedad, y respaldaron la Declaración "*Cooperation for Beneficial AI*", habiéndose publicado el denominado "*Tokio Statement*", en que se puso de manifiesto, respecto del impacto global de la IA y sus retos: "*Instamos a que estos desafíos se aborden con un espíritu de cooperación, no de competencia. Nuestra tarea colectiva debe ser la de garantizar que la IA contribuya al crecimiento humano sostenible en todo el mundo. Debe ser demostrablemente segura, confiable y robusta, y debe desarrollarse en alineación con los valores de las comunidades en las que se desarrollará*."[138]

Merecedora de especial reconocimiento es la creación del AI Now Institute en la Universidad de Nueva York, un centro de investigación multidisciplinar dedicado a explorar y entender las implicaciones sociales de la IA que se está postulando como uno de los grandes centros de referencia en el ámbito de tal tecnología.

De un modo más genérico pero no por ello menos interesante, siendo que resulta plenamente aplicable, en muchos aspectos, en el ámbito de la IA, la Universidad de Deusto presentó la "*Declaración de Derechos Humanos en entornos digitales*"[139], entre los que se hallan el derecho a la protección de la integridad personal ante la tecnología, el derecho a la privacidad en entornos tecnológicos, el derecho a la transparencia y a la responsabilidad en el uso de algoritmos, el derecho a disponer de una última instancia humana en las decisiones de sistemas experto y el derecho a una red segura.

[138] Véase AI, s.f..
[139] Véase Deusto, s.f..

Finalmente, la UNI Global Union, una federación sindical a nivel mundial basada en Nyon (Suiza), que representa a más de 20 millones de trabajadores de más de 150 países y lidera el proyecto "*Future World of Work*", publicó los "*Top 10 Principles for Ethical Artificial Intelligence*", declarando que: "*La IA debe poner a las personas y al planeta por delante. Es por ello que las discusiones éticas sobre IA a escala global son esenciales. Una convención global sobre IA ética que abarque todos los ámbitos es la garantía más viable para la supervivencia humana*."[140]

En el año 2018, OPEN AI, un laboratorio de investigación con base en San Francisco (California, EEUU), creado en 2015 por un grupo de investigadores europeos y estadounidenses con el objetivo de asegurar un uso de la IA beneficioso para la humanidad, tal y como ya se ha anunciado con anterioridad (entre cuyos inversores se hallan Microsoft, Reid Hoffman's charitable Foundation, y Khosla Ventures), publicó la "*OPEN AI Charter*", un documento que sentó sus bases de actuación con el firme compromiso de operar siempre en beneficio colectivo y asegurar que la IA no se emplea para dañar a la humanidad o concentrar indebidamente el poder en manos de unos pocos.[141]

La Association for Computing Machinery US Public Policy Council (USACM), ya en 2017 publicó un "*Statement on Algorithmic Transparency and Accountability*"[142], y en 2018 el Comité de ética profesional de la Association for Computing Machinery's (ACM) Committee on Professional Ethics hizo público su tercer borrador del "*Code of Ethics and Professional Conduct*"[143], diseñado para apoyar a todos los profesionales informáticos, actuales y futuros, instructores, personas influyentes y cualquier otra que utilice la tecnología de modo que cause impacto en la sociedad, y asimismo, establecer las bases para la actuación en caso de vulneraciones.

Asimismo, The Future of Privacy Forum (FPF), una organización sin ánimo de lucro con sede en Washington DC que sirve como catalizadora para el liderazgo y el mecenazgo en el ámbito de la privacidad, con apoyo a las tecnologías emergentes, publicó un documento titulado "*Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning*

[140] UNI Global Union, 2017, pág. 5.
[141] Véase Open AI, 2018.
[142] Véase ACM, 2017.
[143] Véase ACM Ethics, 2018.

Models"[144], con el foco puesto principalmente en los datos que nutren los sistemas de aprendizaje maquinal y la cuestión de la transparencia de estos.

Especial mención merece la publicación de los "*Principles, Policies and Laws for the Responsible Use of AI*" realizada por Microsoft en enero del 2018[145], y la Declaración sobre IA publicada por Sundar Pichai, Consejero Delegado de Google, en junio del mismo año[146], lo que pone de manifiesto la implicación de los grandes gigantes tecnológicos en la creación de una regulación responsable de la IA.

También en la prestigiosa revista Science se publicó un artículo titulado "*How AI can be a force for good*?" [147] , en que se pusieron de manifiesto los posibles riesgos e implicaciones éticas de la IA y se establecieron una serie de bases para un buen uso de la misma. En la introducción de dicho artículo se dispone que: "*La IA no es solo una nueva tecnología que requiere regulación. Es una fuerza poderosa que está cambiando las prácticas diarias, las interacciones personales y profesionales y los entornos. Para el bienestar de la humanidad es crucial que tal poder se use como una fuerza para hacer el bien. La ética juega un papel clave en este proceso, siendo necesario garantizar que las regulaciones de la IA aprovechen su potencial y al mismo tiempo mitiguen sus riesgos*."

The Public Voice Coalition, una organización que promueve la participación pública en las decisiones sobre el futuro de Internet, publicó en una conferencia celebrada en Bruselas (Bélgica) los denominados "*Universal Guidelines for Artificial Intelligence*" [148] que incluye, entre otros, el derecho a la transparencia, la explicabilidad y la publicidad, el derecho a una decisión final tomada por un humano, la obligación de equidad y justicia, la obligación de ciberseguridad, entre otras.

Especialmente interesante es el proyecto "*Algorithms and Justice*" creado por el Berkman Klein Center for Internet & Society de la Universidad de Harvard y el MIT Media Lab, que trata de conseguir que "*las instituciones gubernamentales incorporen IA, algoritmos y*

[144] Véase Future of Privacy Forum, 2018.
[145] Véase Microsoft Corporation, 2018.
[146] Véase Google, 2018.
[147] Taddeo & Floridi, 2018, págs. 751-752.
[148] Véase The Public Voice, 2018.

tecnologías de Machine Learning en su toma de decisiones", y explorar "*vías en las que el desarrollo de tales tecnologías por actores públicos y privados pueda causar impacto en los derechos de los ciudadanos y llegue a conseguirse justicia social*."[149]

En el año 2019, llegó el turno de China de la mano de la publicación de los "*Beijing AI Principles*", avalados por the Beijing Academy of Artificial Intelligence (BAAI), la Universidad de Pekin y de Tsinghua (China), el Institute of Automation and Institute of Computing Technology in Chinese Academy of Sciences, y varias grandes compañías chinas tales como Baidu, Alibaba y Tencent, que de forma principal abogan por una AI alineada con los valores humanos[150]. El documento publicado resultó tener como objetivo llamar *"al desarrollo saludable de la IA para lograr una comunidad de destino común, y la realización de una IA beneficiosa para la humanidad y la naturaleza*."[151] Entre otros, se recoge la necesidad de desarrollo de una IA beneficiosa para la humanidad, responsable, con control de riesgos, ética, diversa pero inclusiva, abierta y compartida, con un uso sujeto al consentimiento informado de los ciudadanos, con educación y formación al respecto, optimización del empleo, harmonía y cooperación y, sobre todo, un plan a largo plazo.

Asimismo, la compañía Ericsson publicó un artículo titulado: "*Ethics and AI: 8 steps to build trust in intelligent technology*"[152], en que se establecieron 8 pasos para construir una IA confiable, basada en metodologías de programas líderes sobre ética y "*compliance*", y se explicó por qué era esencial que las empresas las adoptaran para el uso de la IA.

También en Brasil se creó un nuevo instituto para la investigación de IA: The Advanced Institute for AI (AI2)[153], con el objetivo contribuir al desarrollo de la sociedad mediante la fusión de los intereses y la experiencia del sector académico y el sector privado, promoviendo la colaboración entre ambos.

Por su parte, en 2019 The Law Society of England and Wales elaboró un informe titulado "*Algorithms in the justice system*", en que se ponen de manifiesto los derechos fundamentales que podrían verse vulnerados por el uso de algoritmos en la Administracion

[149] Berkman Klein Center & MIT Media Lab, 2018.
[150] Véase BAAI, 2019.
[151] Journalism AI, 2019.
[152] Véase Ericsson, 2019.
[153] Véase Advanced Institute for Artificial Intelligence, 2019.

de Justicia (derecho a la libertad y seguridad, derecho a un juicio justo, respeto por la vida privada y familiar, derecho a la libertad de expresión, derecho a la libertad de reunión y asociación, derecho a un recurso efectivo y derecho a la no discriminación), y se elabora un mapa del uso de distintas herramientas de IA en los sistemas judiciales del país. [154]

En el año 2020, la Universidad de Harvard (Massachusetts, EEUU) publicó una guía para desarrollar una IA respetuosa con la ética: "*A P* (Ericsson, 2019)*ractical Guide to Building Ethical AI*".[155]

Asimismo, por su parte, el AI Ethics Lab de la Association for Computing Machinery (ACM) creó "*Dynamics of AI Principles*"[156], una caja de herramientas interactiva con características para localizar, ordenar y visualizar conjuntos de principios de IA; comparar puntos clave de diferentes conjuntos de principios; y sistematizar la relación entre principios.

Y la Universidad de Texas (EEUU) celebró el evento 2020 Global Analytics Summit: Ethics in AI[157], con el objetivo de explorar cómo la sociedad y las organizaciones podían maximizar los beneficios y minimizar los riesgos los algoritmos.

No obstante todo lo anterior, y a pesar de las bondades y las ventajas de la gran proliferación mundial de principios que deben regir la IA, en la "*AI, Ethics and Society Conference*", organizada en Honolulu (Hawai, EEUU) por la Association for the Advancement of Artificial Intelligence (AAAI), la Association for Computing Machinery (ACM) y el ACM Special Interest Group on AI, un grupo de investigadores de la Universidad de Cambridge (Reino Unido), puso de manifiesto que "*si bien es importante articular y acordar principios sobre IA, ello es solo un punto de partida*."[158] Así, basándose en comparaciones con el campo de la bioética, destacan algunas de las limitaciones de tales principios: "*en particular, a menudo son demasiado amplios y de un nivel demasiado alto como para guiar a la ética en la práctica. Sugerimos que un próximo paso importante para el campo de la ética de la IA es centrarse en explorar las tensiones que surgen*

[154] Véase The Law Society of England and Wales, 2019.
[155] Véase Blackman, 2020.
[156] Véase AI Ethics Lab, 2020.
[157] Véase University of Texas, 2020.
[158] Whittlestone, Alexandrova, Nyrup & Cave, 2019, pág. 195.

inevitablemente cuando intentamos implementar tales principios en la práctica. Al reconocer explícitamente dichas tensiones, podemos comenzar a tomar decisiones sobre cómo deben resolverse casos específicos, y desarrollar marcos y pautas para el desarrollo ético de la IA que sean rigurosas y con relevancia en la práctica." De acuerdo con ello, se plantearon algunos escenarios en los que surgen tensiones relativas a la ética de la IA y se discutió sobre los procesos que podrían ser necesarios para resolverlas.

Y es que está claro que ha llegado el momento de pasar de la teoría a la acción, puesto que si no, resulta imposible avanzar de forma segura y competitiva. Es evidente que era absolutamente necesaria la fase vivida hasta el momento, puesto que resultaba imprescindible la realización de un ejercicio común de concienciación, compromiso y diseño de estrategias para garantizar un diseño, desarrollo y uso de la IA beneficioso para la humanidad, si bien tal y como se puso de relieve en el referido evento, es hora de ponerse "manos a la obra". Y entiendo que ese va a ser el *alma mater* del nuevo periodo que tenemos por delante, en que, por un lado, van a ir surgiendo millones de casuísticas distintas que habrá que ir resolviendo caso por caso con aplicación de los principios que rigen el uso de la IA en cada territorio; y, por otro lado, de forma reactiva (ya que, por desgracia, la fase de prevención ya ha quedado atrás), irán aprobándose leyes más concretas y específicas que irán guiando a todos los operadores del ámbito de tal tecnología. Esta útima tarea, no obstante, no es fácil, habida cuenta de que existe una línea muy fina entre la regulación, necesaria para garantizar un uso racional de la IA, y la sobrerregulación, que podría implicar importantes dificultades e ineficiencias de toda clase.

2.4.2.2. LA REGULACIÓN EN EL ÁMBITO PÚBILCO

Por otro lado, en el ámbito del sector público y gubernamental, que tal y como se ha advertido con anterioridad ha tenido una intervención más reactiva que proactiva, y sin duda, más lenta y menos intensa que la del sector privado, es importante distinguir entre las iniciativas surgidas a nivel nacional, en los distintos países del mundo, e internacional.

-A NIVEL ESTATAL

A *nivel estatal*, en ***Europa*** es interesante analizar las políticas y estrategias adoptadas por distintos países.

En España, ya en febrero del año 2013 se publicó por el Ministerio de Industria, Energía y Turismo, la denominada "Agenda Digital para España"[159], como marco de referencia para establecer una hoja de ruta en materia de Tecnologías de la Información y las Comunicaciones (TIC) y de Administración electrónica, con el objetivo de establecer la estrategia de España para alcanzar los fines de la Agenda Digital para Europa; maximizar el impacto de las políticas públicas en TIC para mejorar la productividad y la competitividad; y transformar y modernizar la economía y la sociedad española mediante un uso eficaz e intensivo de las TIC por la ciudadanía, las empresas y las Administraciones.

Desde 2015, la Secretaría General de Industria y de la PYME desarrolló el contenido de la estrategia de Industria Conectada 4.0,[160] que respondía a un triple objetivo: incrementar el valor añadido industrial y el empleo cualificado en el sector industrial; favorecer el modelo industrial de futuro para la industria española, con el fin de potenciar los sectores industriales de futuro de la economía y aumentar su potencial de crecimiento, desarrollando a su vez una oferta local de soluciones digitales; y desarrollar palancas competitivas diferenciales para favorecer la industria española e impulsar sus exportaciones[161].

En el año 2015, asimismo, se publicó por la Secretaría de Estado de Telecomunicaciones y la Sociedad de la Información el "Plan de Impulso de las Tecnologías del Lenguaje"[162] con el objetivo de fomentar el desarrollo del procesamiento del lenguaje natural y la traducción automática en lengua española y lenguas cooficiales a través del aumento del número, calidad y disponibilidad de las infraestructuras lingüísticas en español y lenguas cooficiales; el impulso de la industria del lenguaje; y la mejora de la calidad y capacidad del servicio público.

Posteriormente, en el año 2018, la Confederación Española de Organizaciones Empresariales (CEOE) publicó el "Plan Digital 2025: la digitalización de la sociedad española"[163], que planteó un consenso político, económico y social mediante un Acuerdo para la Digitalización, y desarrolló las estrategias y las propuestas que deberían ejecutarse para lograr los objetivos perseguidos, a saber: reforzar la capacidad de coordinación

[159] Véase Ministerio de Transformación Digital, 2013.
[160] Véase Ministerio de Industria, 2015.
[161] Ministerio de Industria, Industria Conectada 4.0, s.f..
[162] Ministerio de Transformación Digital, 2015.
[163] CEOE, 2018.

transversal; la asignación a una Comisión del Congreso de los Diputados de la "Digitalización de España" para seguir el Acuerdo para la Digitalización, garantizar su desarrollo y verificar que cuenta con el marco legal y con los medios adecuados para su ejecución; la coordinación de las medidas impulsadas por el Acuerdo con la Estrategia Europea de Digitalización para el año 2025; el inicio de un diálogo social sobre las repercusiones tecnológicas en el trabajo del futuro y el futuro del trabajo, mediante la creación de una Comisión con el Gobierno, los diversos representantes sociales y los colegios profesionales del ámbito tecnológico; el desarrollo prioritario y con la intensidad necesaria de los tres pilares básicos para la digitalización: educación y formación contínua, innovación y emprendimiento; garantizar que las Administraciones Públicas, a todos los niveles, sean más eficientes y más inteligentes, mediante el cumplimiento efectivo del mandato de lograr una Administración 100% electrónica de la Ley 39/95, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas; acelerar la digitalización de los diferentes sectores productivos; asegurar que todos los ciudadanos puedan estar conectados con acciones en tres ámbitos específicos: derechos digitales, confianza digital y fomento de la demanda; incrementar la utilización de las nuevas tecnologías en beneficio de una sociedad sostenible; y realizar campañas de sensibilización y concienciación sobre la urgencia con la que hay que abordar la digitalización en todos los ámbitos.

En 2019, el Ministerio de Ciencia, Innovación y Universidades publicó la "Estrategia Española de I+D+I en IA"[164] que estableció una serie de prioridades, a saber: lograr una estructura organizativa que permitiera desarrollar un sistema de I+D+I en IA y medir su impacto; establecer áreas estratégicas en las que es necesario centrar los esfuerzos de las actividades de I+D+I; facilitar la transferencia del conocimiento y su retorno a la sociedad; planificar las acciones de formación y profesionalización en el ámbito de la IA; desarrollar un ecosistema digital de datos y valorizar las infraestructuras disponibles; y analizar la ética de la IA desde la perspectiva de la I+D+I. Asimismo, se realizaron las siguientes recomendaciones: lanzar una Estrategia Nacional para la IA que permitiera el desarrollo e implementación de medidas específicas dirigidas a los sectores estratégicos nacionales, previendo que la evaluación y seguimiento de tales medidas fuera realizada a través de un

[164] Ministerio de Ciencia, 2019.

Observatorio Español de la IA; aprovechar la IA para alcanzar los objetivos marcados en la Agenda 2030; diseñar e implementar actuaciones específicas que impulsaran la transferencia de conocimiento al entorno socioeconómico; lanzar o adaptar los programas de fomento de vocaciones, no limitado a la I+D, así como la atracción, retención y recuperación de talento dirigidas a la IA; usar la IA para garantizar un uso óptimo de los datos abiertos; crear un Instituto Nacional de Datos que permitiera planificar y definir una gobernanza sobre los datos procedentes de los diferentes niveles de la Administración Pública; detectar las necesidades de adaptación y mejora de competencias en los distintos niveles de nuestro sistema educativo; y velar por que todas las actividades e iniciativas derivadas de los marcos estratégicos enfocados al desarrollo de la IA, así como sus resultados cumplieran con los compromisos éticos, legales y sociales de nuestro país y de nuestro entorno europeo.

El 23 de julio de 2020 el gobierno presentó la Agenda España Digital 2025, que incluyó casi cincuenta medidas con las que se pretendía impulsar el proceso de transformación digital del país a lo largo de los siguientes cinco años, a través de una colaboración entre el sector público y el privado.[165] Una de sus principales propuestas, que asimismo fue incluida en el Plan de recuperación, transformación y resiliencia de la economía española presentado por el gobierno en octubre de 2020, fue la de la publicación de una Estrategia Nacional de IA[166], que finalmente vio la luz el 2 de diciembre de ese mismo año con la finalidad de "*vertebrar la acción de las distintas administraciones y proporcionar un marco de referencia e impulso para el sector público y privado*"[167] en el ámbito de tal tecnología y responder "*al compromiso compartido con nuestros socios europeos para que la UE se sitúe como líder en esta materia*."[168]

Dicha Estrategia recoge siete objetivos estratégicos[169], a saber:

-conseguir excelencia científica e innovación en IA;
-incrementar la proyección de la lengua española;
-crear empleo cualificado;

[165] Véase Gobierno de España, 2020.
[166] Véase Gobierno de España, La Moncloa, s.f..
[167] Pág. 11.
[168] Pág. 13.
[169] Pág. 15.

-transformar el tejido productivo;
-generar un entorno de confianza en relación a la IA;
-impulsar valores humanistas en la IA; y
-potenciar una IA inlusiva y sostenible.

Y, para dar cumplimiento a los mencionados objetivos, en la Estrategia asimismo se definen seis ejes de actuación que agrupan las acciones prioritarias a llevara a cabo en el período 2020-2025, a saber:

-eje estratégico 1: impulsar la investigación científica, el desarrollo tecnológico y la innovación en IA;
-eje estratégico 2: promover el desarrollo de capacidades digitales, potenciar el talento nacional y atraer talento global;
-eje estratégico 3: desarrollar plataformas de datos e infraestructuras tecnológicas que den soporte a la IA;
-eje estratégico 4: integrar la IA en las cadenas de valor para transformar el tejido económico;
-eje estratégico 5: potenciar el uso de la IA en la Administración Pública y en las misiones estratégicas nacionales;
-eje estratégico 6: establecer un marco ético y normativo que refuerce la protección de los derechos individuales y colectivos, a efectos de garantizar la inclusión y el bienestar social.

Finalmente, se ponen de manifiesto una serie de desafíos a afrontar para poder lograr una IA sostenible e inclusiva, a saber: reducir la brecha de género del ámbito de la IA en empleo y liderazgo; favorecer a la transición ecológica y a la reducción de la huella de carbono; favorecer a la vertebración territorial del país; y favorecer a la reducción de la brecha digital.

En mi opinión, no obstante, la Estrategia Nacional de IA llega tarde, inaceptablemente tarde. Y es que ya en 2017 se anunció[170] por el Ministerio de Industria, Comercio y Turismo que se había nombrado a un grupo de expertos o "sabios" en IA[171] para que

[170] Véase Ministerio de Industria, 2017.
[171] Elena Alfaro, Elena Gil, Asunción Gómez, Lorena Juame-Palasí, Miguel Luengo-Oroz, Nuria Oliver, Andrés Predeño, Javier Plaza y Eduardo Vázquez de Castro.

elaboraran un Libro blanco que nunca llegó a ver la luz (a pesar de que, en principio, iba a publicarse en un plazo máximo de seis meses), y el contenido del mismo es el que, en gran parte, ha dado forma a la mencionada Estrategia que, no obstante, sin justificación alguna, no ha sido presentada hasta finales del año 2020, colocando a España, una vez más, a la cola de la innovación.

En Francia, el 29 de marzo del 2018, durante el coloquio "*AI for Humanity*" que se celebró en el Collège de France, el Presidente de la República, Emmanuel Macron, anunció la Estrategia Nacional de IA, inspirada en el informe elaborado por el Diputado y matemático Cédric Villani, que recibe por título: "*Donner un sens à l'Intelligence Artificielle. Pour une stratégie nationale et européenne*"[172], con asignación de mil quinientos millones de euros para inversión en IA para 2022 con el fin de convertir a Francia en un país líder y referente en la investigación de tal tecnología, y la creación de una agencia central de datos.

El 28 de noviembre del 2018, por su parte, Frédérique Vidal, Ministro de Educación Superior, Investigación e Innovación, y Mounir Mahjoubi, Secretario de Estado para una Francia Digital, presentaron en Toulouse las principales directrices de la Estrategia Nacional de investigación en IA, con una financiación de seiscientos sesenta y cinco millones de euros para 2022, que se basan principalmente en la investigación, clave para el desarrollo de la IA, y apuntan a un doble objetivo: mantener de modo permanente a Francia en el top 5 de países expertos en IA a escala mundial, y hacer de Francia el país líder europeo en investigación de tal tecnología.[173]

En Italia, el 21 de marzo del 2018 se publicó el Libro Blanco sobre IA bajo el título: "*L'Intelligenza Artificialle al servizio del cittadino*"[174], llevado a cabo por el Grupo de Trabajo sobre IA de la Agencia para la Italia Digital del gobierno italiano, estando centrado principalmente en las posibilidades de incremento de eficiencia y satisfacción del usuario que ofrece la IA en el ámbito de las Administraciones públicas y gubernamentales, con identificación de aquellas áreas principales en las que la IA puede servir de ayuda, a saber: sanidad, educación y Administración de Justicia, así como empleo público y seguridad,

[172] Véase Gobierno de Francia, 2018.
[173] Véase Gobierno de Francia, 2018.
[174] Gobierno de Italia, 2018.

incluyendo la creación de un Centro Nacional de Competencia y un Centro Interdisciplinario sobre IA.

Asimismo, en el seno de la propia Agencia para la Italia Digital se creó el denominado "*AI Ecosystem*", con el que se pretende "*rastrear a los productores y usuarios italianos de sistemas de IA (startups, empresas, entidades de investigación públicas y privadas, Administraciones Públicas, etc), para facilitar la construcción de relaciones, el intercambio de conocimientos y permitir a Italia conocer sus puntos fuertes en el campo de la IA.*"[175]

En agosto de 2019, el Ministerio de Desarrollo Económico de Italia lanzó una versión borrador de su Estrategia Nacional de IA para consulta pública, y al mismo tiempo publicó un documento titulado "Propuestas para una estrategia italiana de IA", que establecía unos principios rectores iniciales y una serie de recomendaciones como base para la creación de la estrategia definitiva de IA de Italia, con la mirada puesta en un plan a largo plazo para el desarrollo sostenible de tal tecnología. La consulta pública se cerró el 13 de septiembre del 2019 y el 2 de julio de 2020 se publicó la Estrategia Nacional de IA,[176] que está estructurada en tres partes: la primera analiza los mercados de IA globales, europeos e italianos; la segunda describe los elementos básicos de la Estrategia; y la tercera examina la gobernanza propuesta en materia de IA y proporciona recomendaciones al respecto.

En Alemania, en diciembre del 2018 el Gobierno federal lanzó su Estrategia Nacional de IA[177] con el compromiso de asumir la tarea de proporcionar una respuesta política a los rápidos avances de tal tecnología y hacer un uso de la misma en beneficio de la sociedad. En dicho documento, se puso de manifiesto la voluntad de salvaguardar la destacada posición de Alemania como centro de investigación, desarrollar la competitividad de la industria alemana y promover el uso de la IA en todos los sectores del país, centrándose en los beneficios para las personas y para el medio ambiente, con establecimiento de un diálogo intensivo y constante con todos los sectores implicados.

[175] Gobierno de Italia, 2018.
[176] Véase Gobierno de Italia, 2020.
[177] Gobierno de Alemania, 2018.

En el presupuesto federal de 2019 se dio un primer paso, asignando un total de quinientos millones de euros para reforzar la estrategia de IA para dicho año y los siguientes, y hasta 2025 inclusive, el gobierno alemán tiene la intención de proporcionar alrededor de tres mil millones de euros para la implementación de dicha Estrategia.

Ya en junio del 2017, no obstante, el Ministerio de Transporte e Infraestructuras Digitales, a través de la Comisión Ética del mismo, había publicado una serie de reglas éticas para la toma de decisiones programadas integradas en los vehículos autónomos.[178]

En Reino Unido, el 1 de marzo del 2017 se publicó la Estrategia Digital de dicho país ("*UK Digital Strategy 2017*")[179], que identificó la IA como un elemento clave para el crecimiento de la economía digital inglesa y asignó diecisiete millones de libras a distintos centros universitarios para el desarrollo de tal tecnología y la robótica, habiendo el gobierno incrementado la inversión en investigación de IA en casi cinco mil millones de libras.

El 15 de octubre del 2017 el gobierno de Reino Unido publicó un informe independiente titulado "*Growing the Artificial Intelligence industry in the UK*", elaborado por los profesores Dame Wendy Hall y Jérôme Pesenti, que informaba sobre cómo la industria de la IA podía crecer en el país, identificando a este como un centro internacional de experiencia en el ámbito de tal tecnología, y establecía dieciocho recomendaciones, entre otras, la mejora del acceso a los datos y el intercambio de los mismos a través de "*Data Trusts*", y el desarrollo de un marco para mejorar la transparencia y la rendición de cuentas de las decisiones impulsadas por la IA.[180] Y, tan solo un mes más tarde, el 27 de noviembre del 2017, el gobierno publicó el Libro Blanco "*Industrial Strategy: Building a Britain fit for the future*"[181], en que se dispuso que "*la IA transformará la forma en que vivimos y trabajamos, desde la forma en que diagnosticamos y tratamos el cáncer hasta la seguridad de las transacciones en línea. Esta cuarta revolución industrial es de una escala, velocidad y complejidad sin precedentes*"[182], y calificó tal tecnología como uno de los grandes desafíos para llevar al Reino Unido a la vanguardia de la industria.

[178] Véase Gobierno de Alemania, 2017.
[179] Véase Gobierno de Reino Unido, 2017.
[180] Véase Gobierno de Reino Unido, 2017.
[181] Véase Gobierno de Reino Unido, 2017.
[182] Pág. 32.

Posteriormente, en la misma línea, el 26 de abril del 2018, se publicó la Estrategia Nacional de IA en un documento titulado "*Industrial Strategy: Artificial Intelligence Sector Deal*"[183], que estableció como prioridad la creación de una economía que sacara provecho de las grandes oportunidades de esta era: la IA y el *big data* y anunció, como acciones del gobierno, con soporte financiero (hasta novecientos cincuenta millones de libras), el apoyo a la innovación e investigación de la IA y la implementación del uso de dicha tecnología, incluso en el sector público.

Y en enero del 2020 se publicó una guía para el uso de la IA en el sector público "*A guide to using Artificial Intelligence in the public sector*"[184] , y se anunció la creación de tres nuevos organismos para apoyar el uso de IA, construir la infraestructura adecuada y facilitar la adopción de la misma por parte del sector público y privado: el Consejo de AI (AI Council), la Oficina de IA (AI Office) y el Centro de Ética e Innovación de Datos (Centre for Data Ethics and Innovation), y de dos fondos para apoyar el desarrollo y la adopción de sistemas de IA: el GovTech Catalyst y el Regulators' Pioneer Fund.

En Países Bajos, el 1 de junio del 2018 se publicó la "*Dutch Digitalisation Strategy*"[185], un informe en que se examina qué es necesario para preparar a dicho país para el futuro digital, sobre la idea de que la digitalización está transformando las distintas economías y sociedades del mundo a un ritmo rápido. En relación con ello, se marcaron dos líneas de actuación: por un lado, la del aprovechamiento de las oportunidades económicas y sociales que ofrece la digitalización, entendiendo que los Países Bajos debían liderar el camino con la investigación, los experimentos y la aplicación de las nuevas tecnologías; y por otro lado, la del aumento de la confianza de los ciudadanos y las empresas en las nuevas tecnologías, con fortalecimiento de las bases para la digitalización, especialmente en los ámbitos de la protección de la privacidad, la ciberseguridad, las habilidades digitales y la competencia leal.

El 6 de noviembre del 2018 AI for the Netherlands (AINED), una coalición público-privada de industria y academia, de la cual forma parte la Organización de Investigación Científica de los Países Bajos, publicó un informe para ayudar a los Países Bajos en la

[183] Véase Gobierno de Reino Unido, 2018.
[184] Véase Gobierno de Reino Unido, 2020.
[185] Véase Gobierno de Países Bajos, 2018.

creación de una estrategia nacional de IA y alcanzar éxito en tal campo. En tal documento se establecieron una serie de objetivos y actuaciones, entre los que se preveía la creación de marcos socioeconómicos y éticos para el uso de tal tecnología y el fomento de la cooperación público-privada en sectores clave.[186]

Finalmente, en octubre del 2019 el gobierno de los Países Bajos publicó su Plan estratégico de IA "*Strategisch Actieplan voor Artificiële Intelligentie*" [187] que contiene un conjunto de iniciativas políticas y planes de acción para fortalecer la competitividad de dicho país en el ámbito de la IA en el mercado global basándose principalmente en la necesidad de aprovechar las oportunidades sociales y económicas que brinda tal tecnología; crear las condiciones adecuadas para ello; fomentar la investigación y la innovación en IA, facilitando el acceso a datos de calidad y mejorando la infraestructura digital; y fortalecer los cimientos, incluyendo acciones políticas relacionadas con la ética, la confianza y la seguridad de los ciudadanos y los derechos humanos.[188]

En Dinamarca, en enero del 2018 el Ministerio de Industria, Negocios y Asuntos Financieros publicó la estrategia para el crecimiento digital en Dinamarca "*Strategy for Denmark's Digital Growth*"[189], con la intención de crear las condiciones adecuadas para que las empresas danesas utilizaran las nuevas tecnologías, habida cuenta de los beneficios que la transformación digital podía implicar tanto para el comercio y la industria, como para el individuo y la sociedad en general. Según lo que se desprende de tal iniciativa, la intención del gobierno era que Dinamarca se conviertiera en un líder digital, habiendo comprometido mil millones de DKK (unos ciento cincuenta mil millones de euros) hasta 2025 para implementar la estrategia, con seis áreas principales de enfoque: la creación de un "*Digital Hub*" para fortalecer el crecimiento digital, la mejora digital de las PYMEs, el fomento de habilidades digitales para todos, la consideración de los datos como motor de crecimiento en el comercio y la industria, la regulación ágil de estos dos últimos sectores y el fortalecimiento de la ciberseguridad en las empresas.

[186] Véase AINED, 2018.
[187] Véase Gobierno de Países Bajos, 2019.
[188] OECD, 2019.
[189] Gobierno de Dinamarca, 2018.

Y en marzo del 2019, el gobierno publicó la Estrategia Nacional de IA[190], que mantiene la intención de convertir a Dinamarca en un país líder digital, en concreto en el ámbito de la IA, y resalta la importancia de que el país aproveche las oportunidades que brinda tal tecnología sin que ello implique, no obstante, una pérdida de confianza o seguridad de los ciudadanos, destacando la necesidad de respetar los valores de libertad, seguridad, igualdad y justicia, y de garantizar que los datos empleados por los sistemas de IA sean de calidad. Asimismo, se dispone claramente que "*la tecnología no puede ni debe reemplazar a las personas ni tomar decisiones por nosotros*", y se destaca que el uso de la IA debe ser un complemento de la toma de decisiones humanas, no una sustitución.

Dicha estrategia contiene veinticuatro iniciativas para las cuales el gobierno danés ha destinado aproximadamente nueve millones de euros para el período 2019-2027, (aunque tal dotación fue posteriormente reducida a cinco millones), además de las partidas presupuestarias destinadas a la inversión digital que se van aprobando en las Leyes de finanzas anuales.[191]

En Noruega, ya en 2016 el gobierno publicó el Libro Blanco "*Digital Agenda for Norway-ICT for a simpler everyday life and increased productivity*"[192] con el objetivo de presentar la política del país en relación con la explotación de las tecnologías de la información y la comunicación (TIC), poniendo el foco en el beneficio de la sociedad, siendo la prioridad la de conseguir una Administración Pública eficiente y centrada en el usuario, y la de la creación de valor e inclusión en todos los ámbitos en que estas operan. Posteriormente, el 22 de junio del 2018 la Autoridad de Protección de Datos de Noruega (DPA) publicó el informe "*AI and Privacy*"[193], que desarrolló las aproximaciones jurídicas y tecnológicas descritas en el informe de 2014 "*Big Data: privacy principles under pressure*"[194], y el 31 de agosto del 2018 se hizo público el informe "*Digital2*1"[195] una estrategia creada para promover la creación y el desarrollo de valor científico en diversas áreas importantes de la sociedad, habiéndose identificado la IA como una de ellas.

[190] Gobierno de Dinamarca, 2019.
[191] Comisión Europea, 2019.
[192] Gobierno de Noruega, 2016.
[193] Véase DPA, 2018.
[194] Véase DPA, 2014.
[195] Gobierno de Noruega, 2018.

El 14 de enero del 2020, por su parte, se publicó la Estrategia Nacional de IA "*The Norwegian national strategy for AI*"[196] sobre la creencia de que tal tecnología no solo permitiría realizar tareas de forma cada vez más eficiente, sino también de modos completamente nuevos, siendo la voluntad del gobierno del país la de tomar la iniciativa de un desarrollo y uso de tal tecnología que respete los derechos y libertades de las personas, partiendo de la posición privilegiada de la que parte. En dicho documento se establecen las bases para que la creación y el desarrollo de la IA en Noruega garantice los principios éticos, el respeto a los derechos humanos y la democracia, promoviendo una IA responsable y confiable, con salvaguarda de la integridad y la privacidad del individuo. Para ello, se prevé la supervisión por parte de las autoridades de los sistemas de IA que operen en su ámbito de actuación.

En Suecia, en mayo del 2018, Vinnova, la agencia de innovación del país, publicó el informe "*Artificial intelligence in Swedish business and society. Analysis of development and potential*" [197], que destacaba las oportunidades y los desafíos de la IA en el país, y las capacidades de este para aprovechar todo el potencial de dicha tecnología.

Y, asimismo, el gobierno publicó su Estrategia Nacional en un informe titulado "*National approach to AI*"[198] , que estableció las directrices generales para el uso y el desarrollo de la IA, con el objetivo de incrementar el bienestar social y la competitividad a través de los beneficios de tal tecnología. Con dicho fin, tal y como se desprende de dicho documento, la estrategia sueca puso el foco en las siguientes áreas prioritarias: educación, investigación, innovación, y un marco de referencia e infraestructura.[199]

En Finlandia, el 18 de diciembre del 2017, el Ministerio de Asuntos económicos y Trabajo publicó la estrategia nacional de IA a través del documento "*Finland's Age of AI*"[200] con el objetivo de posicionar al país como líder en tal tecnología, mediante la adopción de una política de *open data* y la creación de las condiciones adecuadas para un desarrollo próspero de la misma, con el foco puesto en el aumento de la competitividad de las

[196] Gobierno de Noruega, 2020.
[197] Véase Vinnova, 2018.
[198] Gobierno de Suecia, 2018.
[199] Comisión Europea, 2018.
[200] Gobierno de Finlandia, 2017.

empresas y la industria, el fomento de servicios públicos de alta calidad y eficiencia, y la garantía del bienestar de la sociedad y sus ciudadanos.

Posteriormente, el 10 de septiembre del 2018 el mismo Ministerio publicó el informe "*Work in the age of AI: Four perspectives on the economy, employment, skills and ethics*"[201], y el 12 de junio del 2019 "*Leading the way into the age of AI: Final report of Finland's Artificial Intelligence Programme 2019*", en el que se acordó dotar, entre otros, al "*AI Business Programme*" con cien millones de euros en un periodo de 4 años, y al "*The Finnish Centre for AI*" (FCAI) con ocho millones de euros para el periodo 2019-2022.[202]

La República Checa, en mayo del 2019 publicó su estrategia nacional de IA en un informe titulado: "*National AI strategy of the Czech Republic*"[203], con el objetivo de mejorar el crecimiento económico y la competitividad del país en el ámbito de tal tecnología mediante la creación de las condiciones políticas favorables para el desarrollo de sistemas de IA responsables y fiables, la transformación digital de las empresas, en particular las PYMEs, y el desarrollo económico de la sociedad en su conjunto basado en las oportunidades y los beneficios de la IA.[204]

Asimismo, en el mismo año se publicó el plan "*Innovation Strategy of the Czech Republic 2019-2030*", que consta de nueve pilares estratégicos, a saber, la financiación y evaluación de I + D, la educación politécnica, la creación de *Start-ups* nacionales y un ambiente de *Spin-off*, la digitalización, la creación de centros de innovación e investigación, la inversión y el marketing inteligente, la protección de la propiedad intelectual, y la creación de infraestructuras inteligentes.

En Hungría, en octubre de 2018, el gobierno creó la Coalition on AI[205], una asociación entre instituciones gubernamentales, académicos y profesionales de las principales empresas de nuevas tecnologías que está llevando a cabo, entre otras iniciativas, la prepración de una Estrategia Nacional de IA para el país, que fue publicada en mayo de 2020 y tiene como pilares fundamentales: la investigación, el desarrollo y la innovación;

[201] Véase Gobierno de Finlandia, 2018.
[202] Comisión Europea, 2019.
[203] Véase Gobierno de la República Checa, 2019.
[204] Comisión Europea, 2019.
[205] McKenna & Olswang, 2018.

la educación, el desarrollo de competencias en el ámbito de la IA y la preparación social; el desarrollo de infraestructuras; y la creación de un marco regulatorio ético.[206]

Casi un año más tarde, el 21 de junio del 2019, el Ministro de Innovación y Tecnología, en la sesión plenaria de la Coalition on AI celebrada en Budapest, anunció la elaboración de un Plan de Acción con el fin de proporcionar de modo definitivo la base para la Estrategia Nacional de IA.

En Estonia, uno de los países más pioneros en la aplicación de la IA, siendo que desde 2017 ha cedido sus vías públicas para la prueba de vehículos autónomos a través de StreetLEGAL[207] , ya en el año 2016, no obstante, el gobierno de Estonia creó un grupo de trabajo para analizar el problema de la rendición de cuentas en algoritmos de *Machine Learning* y la necesidad de legislación de IA, junto con el Ministerio de Asuntos Económicos y Comunicaciones y la Oficina del Gobierno. Posteriormente, el 25 de julio del 2019 el gobierno adoptó la Estrategia Nacional de IA: "*Estonia's national AI strategy 2019-2021*"[208] , que refleja un conjunto de acciones que el ejecutivo tomará para avanzar en la implementación de la IA tanto en el sector público como en el privado, para aumentar la investigación y el desarrollo (I+D), y para desarrollar un marco legal adecuado, anunciando asimismo que una vez al año se presentará al comité gubernamental que supervisa el desarrollo de la sociedad digital, el e-Estonia Council, una valoración general sobre cómo se está aplicando la estrategia.[209]

En Turquía, el Consejo de Investigación Científica y Tecnológica, la agencia líder para la gestión y financiación de la investigación en tal país, ha financiado numerosos proyectos de I+D en IA.[210]

Además, el 20 de agosto de 2021 se publicó la Circular Presidencial sobre la Estrategia Nacional de Inteligencia Artificial 2021-2025[211], redactada por la Presidencia de la Oficina de Transformación Digital y el Ministerio de Industria y Tecnología en cooperación con el sector público, privado, organizaciones no gubernamentales y universidades. En dicha

[206] Véase Gobierno de Hungría, 2020.
[207] Véase Gobierno de Estonia, 2017.
[208] Véase Gobierno de Estonia, 2019.
[209] Comisión Europea, 2019.
[210] Véase OECD, s.f..
[211] Véase Gobierno de Turquía, 2021.

Estrategia se creó el un Comité Directivo con el fin de promover actividades para determinar las prioridades estratégicas, los objetivos, y las medidas oportunas en el ámbito de la IA e implementarlas.

En Grecia, por su parte, que actualmente está desarrollando su Estrategia Nacional de IA, la IA ya fue reconocida como uno de los principales ejes estratégicos de "La Biblia de Transformación Digital (2020-2025)", el informe político que sirve de hoja de ruta para la transformación digital en tal país. Y en tal documento se determinan los objetivos de la futura Estrategia, a saber:

-determinar las condiciones para el desarrollo de la IA, con la creación de un marco de política de datos, y la definición de unos principios éticos para su desarrollo y su uso seguro;
-describir las prioridades para maximizar los beneficios de la IA con el fin de afrontar los desafíos sociales y el crecimiento económico; y
-analizar cuáles son las acciones neesarias para lograr las prioridades anteriormente mencionadas.

En Malta, el 21 de marzo del 2019 el gobierno presentó para consulta pública el documento "*Malta Towards an AI Strategy*", y en octubre de 2019 publicó el informe titulado "*A Strategy and Vision for Artificial Intelligence in Malta 2030*"[212] con el objetivo de trazar el camino para que dicho país obtenga una ventaja competitiva estratégica en la economía global como líder en el ámbito de la IA.

La Estrategia es expansiva, y analiza el impacto comercial y social, las áreas de oportunidad económica y la necesidad de una consideración especial, si no de una regulación, donde los casos de uso de la IA se cruzan potencialmente con las prioridades, los valores y los derechos de los ciudadanos nacionales.

Finalmente, en Rusia, el Presidente Vladimir Putin introdujo la economía digital como una nueva herramienta para el desarrollo del gobierno, la economía, los negocios y la sociedad

[212] Véase Gobierno de Malta, 2019.

mediante el programa "*Digital Economy of the Russian Federation*"[213], aprobado el 28 de junio de 2017 por la resolución nº1632-r del gobierno ruso.

El 7 de mayo del 2018, en virtud del Decreto del Presidente nº204, se aprobó un Programa de Metas Nacionales y Objetivos Estratégicos del país para 2024, con el fin de lograr avances en ciencia, tecnología (entre otras, IA y robótica) y desarrollo socioeconómico, aumentar la población del país, mejorar el nivel de vida y las condiciones de los ciudadanos, y crear un entorno de oportunidades. [214]

El 10 de octubre del 2019 mediante el Decreto del Presidente nº490 se adoptó la Estrategia Nacional para el Desarrollo de la IA en la Federación de Rusia[215], con el objetivo de sentar las bases para el desarrollo y la mejora de los programas y proyectos del ámbito público y privado que apoyan el desarrollo de IA en dicho país. Los objetivos, entre otros, fueron la investigación en el ámbito de los algoritmos y los métodos matemáticos, el desarrollo de *software* para IA, la recopilación, el almacenamiento y el procesamiento de datos para I+D, así como el aumento de la disponibilidad de un *software* especializado, la mejora de la capacitación del personal en el ámbito de la IA, y el desarrollo de una regulación para el ecosistema de tal tecnología.

Y el 30 de mayo del 2019, el Presidente de la Federación, tras una reunión sobre el desarrollo de la IA, aprobó una serie de instrucciones, una de las cuales, la instrucción 1030, fue dirigida al programa nacional "*Digital Economy of the Russian Federation*", en aras de poner en marcha un proyecto federal destinado a la implementación de la Estrategia Nacional para el desarrollo tecnológico en el campo de la IA de la Federación rusa, incluyendo un plan de acción de tres años.[216]

En ***Asia***, numerosos países han elaborado ya planes o estrategias de IA a nivel nacional, entre otros, los siguientes.

[213] Véase Gobierno de Rusia, 2017.
[214] Gobierno de Rusia, 2018.
[215] Véase Gobierno de Rusia, 2019.
[216] *Idem.*

En India, el Ministro de Finanzas, en su discurso sobre el presupuesto para el ejercicio 2018-2019, ordenó al National Institution for Transforming India un *think tank* formado por grupo de expertos del gobierno indio, que estableciera un Programa Nacional sobre IA, con el objetivo de guiar la investigación y el desarrollo de las tecnologías emergentes. En virtud de ello, dicho organismo adoptó una triple iniciativa: emprender proyectos de investigación de IA en varios ámbitos; elaborar una Estrategia Nacional para construir un ecosistema de IA en la India, para lo cual el 4 de junio del 2018 publicó en su página web el informe titulado: "*National Strategy for AI. #AIForAll*" [217] que se centró en el aprovechamiento de la IA para llevar a cabo un crecimiento inclusivo, mediante la creación y el desarrollo de un ecosistema de investigación, así como el control de los desafíos que tal tecnología presenta, entre otros, las cuestiones relativas a la ética y la privacidad ; y colaborar con expertos y partes interesadas, habiéndose asociado con varios líderes en tecnología de IA para implementar proyectos de tal tecnología en áreas tales como la agricultura y la salud.

En Singapur, en mayo de 2017 el gobierno aprobó un programa denominado "*AI Singapore*" para tratar de sacar el máximo provecho de las posibilidades que ofrece dicha tecnología, con previsión de inversión en IA de ciento cincuenta millones de dólares de Singapur durante los siguientes 5 años. Y, posteriormente, "*AI Singapore*" lanzó dos programas: "*AI for Everyone*" (AI4E) y "*AI for Industry*" (AI4I).[218]

En noviembre del 2017 la Monetary Authority of Singapore (MAS) lanzó una consulta sobre los posibles riesgos éticos del uso de la IA en la industria financiera, con el objetivo de establecer unas pautas regulatorias para el uso ético, responsable y transparente de la IA y el análisis de datos.[219]

Posteriormente, en junio de 2018, el gobierno anunció la creación de un Consejo Asesor de ética, el denominado AI Ethics Council,[220] encabezado por el ex Fiscal General V.K. Rajah y liderado por la Autoridad de Desarrollo de Medios de Infocomm (IMDA), para

[217] Gobierno de India, 2018.
[218] Future of Life Institute, 2019.
[219] Lee, 2017.
[220] Lin, 2018.

asesorar a su gobierno sobre el desarrollo y uso de tal tecnología y trabajar con los comités de ética de las empresas.

El 21 de mayo del 2018 la mencionada Autoridad de Infocomm publicó el "*Digital Economy Framework for Action*", un plan que establece un marco de acción para transformar Singapur en una economía digital líder, identificando la IA como una tecnología fundamental para conseguir dicho objetivo. También en junio de 2018, el Ministro de Comunicaciones e Información del país, calificó la IA como una de las bases sobre las que Singapur planeaba hacer crecer su economía digital. [221]

El 23 de enero del 2019, en la reunión anual del Foro Económico Mundial celebrada en Davos (Suiza), la Comisión de Protección de Datos Personales publicó la primera edición de un modelo de gobernanza de IA "*the Model AI Governance Framework (First Edition)*" para promover la adopción responsable de la IA en Singapur, proporcionando una guía práctica para convertir los principios éticos en prácticas implementables. El 21 de enero del 2020, en el mismo foro, se presentó la segunda edición.[222]

Y, finalmente, el 13 de noviembre del 2019, el Ministro de Asuntos Exteriores y Ministro en funciones de Iniciativa para una Nación Inteligente, anunció en el discurso pronunciado en el Festival Fintech de Singapur, el lanzamiento de la Estrategia Nacional de IA "*Singapore's National AI Strategy*"[223] y manifestó: "*Singapur está listo para desplegar IA a escala nacional: hemos comprometido más de quinientos millones de dólares para financiar actividades relacionadas con IA bajo el Plan de Investigación, Innovación y Empresa 2020. Crearemos una nueva Oficina Nacional de Inteligencia Artificial para unir los esfuerzos de Singapur, ya que nuestro objetivo es ser un líder en el desarrollo y despliegue de soluciones de tal tecnología para 2030. Iniciaremos este esfuerzo al embarcarnos en cinco proyectos nacionales, mientras construimos los facilitadores que sustentan Un vibrante ecosistema de IA*."[224]

En Arabia Saudí, ya en 2016 se anunció el programa "*Vision 2030*", que estableció un plan de reforma económica para estimular el crecimiento de nuevas industrias y diversificar la

[221] Future of Life Institute, 2019.
[222] Véase Gobierno de Singapur, 2020.
[223] Véase Gobierno de Singapur, 2019.
[224] Gobierno de Singapur, 2019.

economía, otorgando a la transformación digital un papel clave para conseguir dicho objetivo mediante el aprovechamiento del *big data*, de la IA y de la automatización industrial.[225]

El 30 de agosto del 2019, además, dicho país anunció la creación de una agencia gubernamental denominada Saudi Data and Artificial Intelligence (SDAIA), con tres objetivos principales: desarrollar estrategias nacionales de datos y de IA; supervisar la ejecución de dichas estrategias nacionales; y crear conciencia sobre tales cuestiones, con comunicación de los logros obtenidos a nivel local y global.

Y el 21 de octubre de 2021 finalmente se publicó su Estrategia Nacional de IA bajo el título "*The Saudi National Strategy for Data & AI*"[226] en el congreso Global AI Summit celebrado en Riyadh, con el objetivo de implementar un plan de múltiples fases que incluye la creación de nuevas políticas, la aprobación de un marco regulatorio completo, y la inversión, la investigación y la innovación en materia de IA.

En Corea del Sur, en marzo del 2016 el gobierno publicó el informe "*Mid-to Long-Term Master Plan in preparation for the Intelligent Information Industry Development Strategy*"[227], que consideraba como fundamental el papel de la IA junto con otras tecnologías emergentes, y destacaba la necesidad de fomentar una sociedad de la información inteligente, sobre la base de una unión de fuerzas público-privada, el diseño y la implementación de políticas equilibradas, que potenciaran las tecnologías y la industria con el fin de alcanzar una sociedad más humana, y el apoyo estratégico para garantizar la seguridad de los derechos.

Y el 17 de diciembre del 2019, el gobierno surcoreano anunció la Estrategia Nacional de IA con el objetivo de impulsar la economía y mejorar el nivel de vida de los ciudadanos para el año 2030[228], a través del impulso del uso de IA por parte de las empresas, con modificación o supresión de las regulaciones excesivamente estrictas y farragosas para lograr la creación de un entorno más favorable para el desarrollo y uso de tal tecnología,

[225] OECD, 2019.
[226] Véase Gobierno de Arabia Saudí, 2021.
[227] Gobierno de Corea del Sur, 2016.
[228] Yonhap, 2019.

identificando como punto fuerte el dominio en el suministro global de *chips* de memoria, anunciando una inversión de casi ochocientos sesenta millones de dólares en la próxima generación de *chips* inteligentes para 2030.[229]

En Japón, en 2016 se creó el "Consejo de Estrategia de Tecnología de IA", que en marzo del 2017 elaboró la "Estrategia de Tecnología de IA", que se centró en promover el desarrollo de tal tecnología.

En marzo del 2017 dicho Consejo de Estrategia de Tecnología de IA publicó una estrategia de tal tecnología que identificó cuestiones críticas relacionadas con esta tales como "*la necesidad de aumentar la inversión, facilitar el uso y el acceso a los datos, y aumentar el número de investigadores e ingenieros de IA*".[230] Posteriormente, el 28 de julio del mismo año, se publicó el "*Draft AI R&D GUIDELINES for International Discussions*", como preparación para la denominada "*The Conference toward AI Network Society*", en relación con actividades de I+D destinadas a promover los beneficios y reducir los riesgos de la IA. Se establecieron nueve principios: de colaboración, de transparencia, de controlabilidad, de seguridad, de privacidad, de ética (respeto de la dignidad humana y de la autonomía individual), de asistencia al usuario, y de responsabilidad.[231]

La Estrategia de Innovación Integrada de Japón, publicada por la Oficina del Gabinete en junio de 2018, contenía un conjunto de acciones de política de IA. La estrategia incluyó la convocatoria de debates de múltiples partes interesadas sobre cuestiones éticas, legales y sociales de la IA, lo que dio lugar a que la Oficina del Gabinete publicara unos Principios sociales para la IA "centrada en el ser humano" en abril de 2019.

En 2019 el gobierno japonés publicó la Estrategia Nacional de IA con el objetivo de incrementar las habilidades para la investigación y la innovación, crear políticas de coordinación vertical y horizontal, fomentar la digitalización, la innovación empresarial y el emprendimiento, afrontar los desafíos sociales y aumentar las capacidades de investigación pública.[232]

[229] Véase Gobierno de Corea, 2019.
[230] OECD, 2019.
[231] Future of Life Institute, 2019.
[232] OECD, 2019.

En China, en 2015 se lanzó el ambicioso plan estratégico "*Made in China 2025*", que causó cierto nerviosismo en el resto del mundo, si bien, como veremos, en 2017 el Consejo de Estado de tal país aprobó un plan de desarrollo para convertirse en el líder mundial de IA en 2030[233], habiendo puesto de manifiesto, en tal caso, la necesidad de crear leyes y regulaciones de IA, así como de establecer unos principios éticos.

En 2016 el gobierno publicó un Plan Nacional de IA, con una duración de tres años, denominado "*Three-year Guidance for Internet Plus Artificial Intelligence Plan (2016-2018)*" creado conjuntamente por la Comisión Nacional de Desarrollo y Reforma, el Ministerio de Ciencia y Tecnología, el Ministerio de Industria y Tecnología de la Información y la Administración del Ciberespacio del país centrado en mejorar la capacidad del *hardware* de IA, crear ecosistemas fuertes, fomentar la aplicación de IA en áreas socioeconómicas importantes y analizar el impacto de la IA en la sociedad, con la creación de un mercado de quince mil millones de dólares para 2018, invirtiendo en investigación y apoyando el desarrollo de la industria china de IA.[234]

Tal y como se ha anunciado, en julio de 2017, el Consejo de Estado de China lanzó el "*Plan de desarrollo de IA de nueva generación*" (AIDP), que describía la estrategia de China para construir una industria nacional de IA por un valor de casi ciento cincuenta mil millones de dólares en los próximos años y para convertirse así en la principal potencia de tal tecnología para 2030, marcando oficialmente el desarrollo del sector como una prioridad nacional.[235] En relación con ello, el Consejo de Estado de China estableció como objetivo para la "tecnología de la información de nueva generación" convertirse en una industria estratégica orientada a representar el 15% del PIB para 2020.[236]

El 18 de enero de 2018, China estableció un Grupo Nacional de Estandarización de IA y un Grupo Asesor nacional formado por expertos en tal tecnología. Al mismo tiempo, el Standardisation Management Committee Second Ministry of Industry publicó un Libro Blanco sobre la normalización de la IA, que fue apoyado por el China Electronic

[233] Mozur, 2020.
[234] OECD, 2019.
[235] Future of Life Institute, 2019.
[236] OECD, 2019.

Standardisation Institute, una división del Ministerio de Industria y Tecnología de la Información. [237]

Además de las iniciativas a nivel central, también diversos gobiernos locales chinos han aprobado interesantes planes relacionados con la IA en los últimos años[238]. Así, por ejemplo, el gobierno de Shanghai, en noviemrbe del 2017, con el objetivo de convertirse en el "*Hub AI*" del país, emitió su propio plan de implementación de IA de nueva generación, con un plan de expansión de la industria en la ciudad a más de veinte mil millones de dólares para 2020[239]; Beijing anunció en 2018 la creación en el distrito de Mentougou de un nuevo e importante parque industrial centrado en la IA[240]; y Guangzhou creó un Instituto Internacional de IA.[241]

En el ámbito de ***Oceanía*** también debe hacerse mención a la aprobación de diversas regulaciones sobre IA a nivel nacional.

En Australia, en 2017 el gobierno publicó un informe titulado "*Australia 2030: Prosperity through innovation. A plan for Australia to thrive in the global innovation race*"[242], con el objetivo de colocar al país, en 2030, a la vanguardia de la carrera de innovación global, basándose en la fuerte y robusta economía ya existente y planenado la conversión de Australia en uno de los mejores lugares del mundo para emprender en innovación, ciencia e investigación y maximizar así los beneficios de todos los australianos.

El 7 de marzo del 2018 se anunció en el Parlamento la creación del grupo parlamentario The Victorian All-Party Parliamentary Group on Artificial Intelligence (VAPPGAI)[243], con el objetivo de discutir sobre la naturaleza compleja y transformadora de la IA desde todos los ángulos del espectro político.

[237] OECD, 2019.
[238] Future of Life Institute, 2019.
[239] The Straits Times, 2017.
[240] The Straits Times, 2018.
[241] You & Berney, 2017.
[242] Véase Gobierno de Australia, 2017.
[243] Véase Gobierno de Australia, 2018.

Asimismo, en mayo del 2018 el Dr. Alan Finkel (científico jefe de Australia) propuso la expedición de certificaciones de IA bajo la denominación "*The Turing Test*" [244] como "sello de confianza" para identificar a aquellas empresas que cumplieran con los estándares éticos y los requisitos de auditoría independiente para el desarrollo de tal tecnología.

En diciembre de 2018 el gobierno lanzó su estrategia de economía digital, "*Australia's Tech Future*"[245], que incluía un análisis sobre el modo en que las empresas, el gobierno y la comunidad podían trabajar de forma conjunta para maximizar los beneficios y las oportunidades que ofrecía la tecnología digital avanzada (entre las que incluía la IA), centrándose en las cuatro áreas clave de personas, servicios, activos digitales y entorno adecuado, con identificación de las prioridades y las acciones gubernamentales clave para cada una de ellas.[246]

El gobierno, además, en su presupuesto para el ejercicio 2018-2019, destinó casi treinta millones de dólares australianos para aumentar los esfuerzos del país en el ámbito de la IA y, en concreto, del *Machine Learning*, a lo largo de los siguientes cuatro años [247]. Dentro de tal plan se incluyó, además, el desarrollo de un Marco de Ética de IA nacional, una hoja de ruta tecnológica y un conjunto de principios.

En 2019, la *Australia's Commonwealth Scientific and Industrial Research Organization* (CSIRO) publicó un documento de debate sobre el marco ético de IA de Australia ("*Australia's Ethics AI Framework*") y lanzó una consulta pública al respecto.[248]

Y en junio de 2021 el gobierno australiano publicó el plan de acción de IA (IA) de Australia ("*Australia's AI Action Plan*"),[249] con el objetivo de que Australia se convierta en un líder mundial en el desarrollo y la adopción de IA confiable, segura y responsable. Con tal finalidad, se incluyen diversas acciones necesarias para garantizar que todos los ciudadanos australianos obtengan los beneficios de la IA, entre otras:

[244] Véase Clark, 2018.
[245] Véase Gobierno de Australia, 2018.
[246] OECD, 2019.
[247] Pearce, 2018.l
[248] Future of Life Institute, 2019.
[249] Véase Gobierno de Australia, 2021.

-impulsar el desarrollo y la adopción de la IA para crear puestos de trabajo e impulsar la productividad;
-crecer y atraer talento y experiencia de otros países del mundo;
-aprovechar sus capacidades para resolver desafíos nacionales y beneficiar a todos los australianos; y
-garantizar que las tecnologías de IA sean responsables, inclusivas y reflejen los valores australianos.

En Nueva Zelanda, en 2017, con el fin de conectar al gobierno con los ciudadanos, las empresas y la academia en el ámbito de la IA, se creó el "*AI Forum of New Zeland*"[250], y en marzo del 2018, se publicó el informe "*Artificial Intelligence: Shaping a Future New Zealand*"[251], que fue la culminación de un proyecto de investigación dirigido por el mencionado AI Forum of New Zeland, que analizó el estado de la industria de IA a nivel nacional e internacional, así como los posibles impactos de tal tecnología en la economía y la sociedad del país, e identificó las oportunidades clave que brinda esta tanto en el sector público como en el privado. Tal documento, si bien no es una estrategia nacional de IA, explora el panorama de dicha tecnología en Nueva Zelanda y los posibles impactos en la economía y la sociedad.

Asimismo, en octubre del 2018 el gobierno publicó "*The Algorithm Assessment Report*", un informe focalizado en el uso de algoritmos en las agencias gubernamentales del país, que incluyó algunos "Principios para el uso seguro y efectivo de datos".[252]

En el ámbito de ***América***, procede establecer una distinción entre América del Norte y América Latina y hacer referencia a algunas de sus iniciativas.

Así, en América del Norte, Canadá, en marzo del 2017 fue el primer país en aprobar un plan estratégico de IA, el denominado "*Pan-Canadian AI Strategy*"-con una dotación de ciento veinticinco millones de dólares canadienses otorgada por el Gobierno del país-, liderada por The Canadian Institute for Advanced Research (CIFAR) en asociación con

[250] Gobierno de Nueva Zelanda, 2017.
[251] Véase Gobierno de Nueva Zelanda, 2018.
[252] Véase Gobierno de Nueva Zelanda, 2018.

tres institutos de IA: Alberta Machine Intelligence Institute (Amii), de Edmonton, Mila, de Monreal, y The Vector Institute, de Toronto, con la finalidad de alcanzar cuatro objetivos principales: lograr un incremento de investigadores y personal cualificado en IA; establecer nodos interconectados de excelencia científica en los tres centros principales de IA del país (Edmonton, Monreal y Toronto); desarrollar liderazgo en el pensamiento global sobre las implicaciones económicas, éticas, políticas y legales de los avances de tal tecnología; y apoyar la creación de una comunidad nacional de investigación sobre IA.[253]

En virtud de ello, Canadá se convirtió en el primer país en adoptar una Estrategia Nacional de IA, con una clara declaración de intenciones: la de convertirse en líder mundial en el campo de tal tecnología.

El 7 de junio del 2019 el Primer Ministro de Canadá y el Presidente de la República Francesa anunciaron la voluntad de ambos países de promover un desarrollo de la IA centrado en el ser humano y crear un "*International Panel on Artificial Intelligence*" (IPAI) para apoyar y guiar un uso responsable de tal tecnología, con respeto y fomento de los derechos humanos, la inclusión, la diversidad, la innovación y el crecimiento económico. En relación con ello, se puso de manifiesto que dicho panel debía servir para facilitar la colaboración internacional entre las múltiples partes interesadas, a saber, la comunidad científica, la industria, la sociedad civil, las organizaciones internacionales relacionadas y los gobiernos.[254]

En 2019, además, el gobierno federal y el de Quebec anunciaron la creación de un centro internacional de expertos en IA con sede en Monreal para fomentar el desarrollo de tal tecnología, como parte de la "*Global Partnership on AI*" (GPAI), con una financiación del gobierno federal de hasta diez millones de dólares canadienses durante cinco años para apoyar las actividades de dicho centro y una subvención de 5 millones de dólares canadienses comprometidos por el gobierno de Quebec.[255]

En Estados Unidos, en mayo del 2017 el Congreso lanzó The bipartisan AI Caucus, que reúne expertos del mundo académico, el gobierno y el sector privado para informar y

[253] Gobierno de Canadá, 2017.
[254] Gobierno de Canadá, 2018.
[255] Kirkwood, 2019.

discutir sobre las implicaciones de tal tecnología. El 7 de noviembre de 2017, dicho Caucus organizó una reunión con The Software & Information Industry Association e IEEE-USA para debatir cuestiones de ética y privacidad relacionadas con la IA bajo el título: "*Machines That Learn: Can They Also Be Taught Human Values?*".

En julio de 2017, el Departamento de Seguridad Nacional publicó un informe titulado "*Artificial Intelligence Risk to Critical Infrastructure*"[256] , que analizó las características de tal tecnología con el fin de obtener una mejor percepción de los beneficios y las amenazas que presenta.

En septiembre del 2017 el Congreso aprobó la denominada ley "*The SELF DRIVE Act*" que requería que el Departamento de Transporte realizara investigaciones sobre la mejor manera de informar a los consumidores de las capacidades y limitaciones de vehículos altamente automatizados.

El 18 de enero del 2018 se presentó ante el Congreso la ley "*The AI JOBS Act 2018*", que estableció: "*Es la sensación del Congreso que la tecnología puede mejorar la vida de las personas, pero también puede perjudicar los trabajos, y por ello, se debe alentar la innovación pero también capacitar y reentrenar a los trabajadores estadounidenses para nuestra economía del siglo XXI*". [257]

En mayo de 2018, la Casa Blanca organizó "*the Summit on AI for American Industry*", que reunió a más de cien altos funcionarios gubernamentales, expertos de las principales instituciones académicas, jefes de laboratorios de investigación industrial y líderes empresariales estadounidenses que estaban trabajando con IA para beneficiar a sus clientes, trabajadores y accionistas, para discutir sobre tal tecnología y las políticas necesarias para implementarla en beneficio del pueblo estadounidense, manteniendo el liderazgo del país.

En noviembre del 2018, para mejorar la coordinación de los esfuerzos federales relacionados con la IA, la Casa Blanca organizó el NSTC Select Committee on AI, una comisión de expertos formada por los más altos funcionarios de I+D del gobierno federal,

[256] Véase Gobierno de Estados Unidos, 2017.
[257] Future of Life Institute, 2019.

con el ánimo de establecer un enfoque conjunto para la planificación y la coordinación de I+D en el ámbito de la IA y el asesoraramiento a la Casa Blanca en tal sentido.[258]

El 11 de febrero de 2019, el presidente Donald Trump firmó la Orden Ejecutiva 13859 que anunciaba la Iniciativa de IA estadounidense: "*The American AI Iniciative*", con el objetivo de promover la IA en EEUU sin dejar de lado los principios que han de regir tal tecnología, tal y como se desprende del capítulo titulado "*AI with American Values*", en el que se establece: "*Nuestro objetivo es asegurar que las tecnologías de IA son comprensibles, de confianza, robustas y seguras*."[259]. A los efectos de conseguir tales fines, se establece una colaboración del gobierno federal con el sector privado, la academia, la sociedad y los socios internacionales con ideas afines, dirigiendo a dicho gobierno a perseguir cinco objetivos para lograr el avance de la IA: invertir en I+D de tal tecnología, liberar recursos, eliminar barreras para la innovación, formar en IA a la mano de obra, y promover un ambiente internacional que apoye la innovación de IA estadounidense y garantice su uso responsable.

En el Congreso, en el mismo mes, se presentó la "*House Resolution 153*" (HRES 153) con el fin de fomentar la creación de pautas para llevar a cabo un desarrollo ético de la IA, con los siguientes objetivos: establecer colaboración entre la industria, el gobierno, la academia y la sociedad civil; fomentar la transparencia y la explicabilidad de los sistemas, los procesos y las implicaciones de la IA; ayudar a empoderar a las mujeres y las poblaciones infrarepresentadas o marginadas; fomentar la privacidad de la información y la protección de los datos personales; incrementar las oportunidades para encontrar salidas profesionales; garantizar la responsabilidad y la supervisión de los sistemas en los casos de toma de decisiones automatizada; fomentar el aprendizaje en STEM ("*Science, Technology, Engineering and Mathematics*"), ciencias sociales y humanidades; incrementar la igualdad en el acceso los servicios y beneficios tecnológicos; potenciar una investigación interdisciplinaria sobre IA segura y beneficiosa; y fomentar la seguridad, la protección y el control de los sistemas de IA ahora y en el futuro.[260]

[258] Gobierno de Estados Unidos, 2018.
[259] Gobierno de Estados Unidos, 2019.
[260] Future of Life Institute, 2019.

El 19 de marzo de 2019, el gobierno federal lanzó la página web AI.gov "*AI for the American People*" para facilitar el acceso de los ciudadanos a todas las iniciativas gubernamentales de IA en curso[261], y en el mismo mes, se creó The bipartisan Senate AI Caucus, que vino a complementar "*The American AI Initiative*" lanzada por la Casa Blanca en febrero ese mismo año, tal y como se ha expuesto con anterioridad.

Asimismo, en junio del 2019 el NSTC Select Committee on AI publicó "*The National AI Research and Development Strategic Plan: 2019 update*"[262] , con el objetivo de guiar al país en las inversiones en I+D en IA, identificando las áreas clave que requerían inversiones federales.

En septiembre de 2019, la Casa Blanca organizó "*The Summit on AI in Government*", en que se reunieron más de ciento setenta y cinco líderes y expertos del gobierno, la industria y la academia con el objetivo de generar y discutir ideas sobre cómo el gobierno federal podía emplear la IA para cumplir su misión de un modo más efectivo y para conseguir mejorar los servicios para el pueblo estadounidense.

En febrero de 2020, la Casa Blanca publicó el informe "*the American AI Initiative: Year One Annual Report*"[263], que puso de relieve los logros conseguidos durante el primer año de vigencia de la Orden Ejecutiva sobre IA, entre otros: la inversión récord en I+D de IA, el desarrollo de la primera Declaración Internacional sobre Principios de IA, y el lanzamiento del primer documento regulatorio de dicha tecnología para garantizar un desarrollo, prueba, implementación y adopción confiables de aplicaciones de IA.

En los presupuestos federales para el año 2020, se llegó a establecer que el ambiente de innovación estadounidense era la envidia del mundo, habida cuenta de que el I+D federal contribuía al crecimiento del empleo, la seguridad nacional y la prosperidad contínua, y se destinaron ciento treinta y cuatro mil millones de dólares para I+D federal.[264]

[261] Future of Life Institute, 2019.
[262] Gobierno de EEUU, 2019.
[263] Gobierno de Estados Unidos, 2020.
[264] Gobierno de Estados Unidos, 2020.

En enero del 2020, la Casa Blanca propuso unos principios reguladores de IA ("*AI Regulatory Principles*")[265] para el uso de tal tecnología en el sector privado.

En el ámbito de la defensa, el 18 de diciembre de 2017, el Presidente Trump firmó una nueva Estrategia de Seguridad Nacional ("*National Security Strategy*")[266] que instaba a Estados Unidos a liderar la investigación y la innovación en tecnologías emergentes, incluyendo la IA. En junio de 2018, el Departamento de Defensa creó el Joint AI Center (JAIC), como centro de referencia en el uso de la IA en misiones clave de defensa. En febrero de 2019, asimismo, el Departamento de Defensa lanzó su Estrategia de IA "*DoD AI Strategy*"[267], que tenía como objetivo el aprovechamiento de las oportunidades que brinda tal tecnología para avanzar en la seguridad y la prosperidad del país. Y, por su parte, en la United States Intelligence Community (IC), el Director Nacional de Inteligencia lanzó en enero de 2019 "*The AIM Initiative. A Strategy for Augmenting Intelligence Using Machines*"[268] , que definió la IA como elemento transformador clave y crucial para el éxito y la eficiencia de la misión de defensa futura, y describió el modo en que el IC tenía planeado aprovechar las capacidades de la IA con el fin de resolver los desafíos legales, políticos, culturales, técnicos y estructurales clave.

Debe ponerse de manifiesto que EEUU, junto con China e Israel, es uno de los países que sin duda se halla a la vanguardia de la IA, habiendo crecido la inversión del gobierno federal en I+D no clasificada para IA y tecnologías relacionadas en más del 40% desde el año 2015[269], lo cual ha implicado enormes beneficios pero también sonadas pugnas de poder con aquellos países que amenazan su soberanía.

En el ámbito de America Latina, varios son los países que han adoptado regulaciones y prometedoras iniciativas en el ámbito de la IA.

En México, ya en el año 1994, la Universidad Veracruzana creó el Centro de Investigación en IA (CIIA), un proyecto innovador en muchos sentidos, habida cuenta de que se puso en marcha a través de un acuerdo de cooperación entre dicha universidad y el Laboratorio

[265] Véase Gobierno de EEUU, 2020.
[266] Véase Gobierno de Estados Unidos, 2017.
[267] Véase Gobierno de Estados Unidos, 2019.
[268] Gobierno de Estados Unidos, 2019.
[269] OECD, 2019.

Nacional de Informática Avanzada (Lania A.C.), lo que permitió combinar los beneficios propios de un laboratorio privado con los de una gran universidad pública. [270]

En 2013, el Gobierno de la República lanzó la Estrategia Digital Nacional, un plan de acción para construir un México Digital, con el objetivo de que la tecnología y la innovación contribuyeran a alcanzar las grandes metas de desarrollo del país.[271] Tras cinco años de implementación, en 2018, con el fin de dar a conocer las mejores prácticas, resultados y metodologías aplicadas con éxito por la mencionada Estrategia entre 2013 y 2018, se pupublicó la "*Estrategia Digital Nacional: el inicio de la transformación digital de México*".[272]

Posteriormente, el 21 de marzo del 2018 se llevó a cabo el anuncio del informe "*En miras hacia una Estrategia de Inteligencia Artificial (IA) en México: Aprovechando la Revolución de IA*" [273] realizado por Oxford Insights y C-Minds, comisionado por la Embajada Británica en México, y se publicó la Estrategia Nacional de México (Estrategia IA-MX 2018) [274], habiéndose convertido tal país en uno de los primeros del mundo en contar con un plan claro y determinado para impulsar la implementación, el desarrollo y uso de la IA.

Dicho plan contempla, entre otros objetivos, el desarrollo de un marco de gobernanza adecuado para fomentar el diálogo multisectorial, a través de la creación de una Subcomisión de IA dentro de la Comisión Intersecretarial para el Desarrollo del Gobierno Electrónico; el mapeo de los usos y necesidades en la industria y la identificación de las mejores prácticas; el impulso del liderazgo internacional de México en la materia, con especial énfasis en la OCDE y el G7; y el trabajo con expertos y ciudadanos mediante la Subcomisión de IA para alcanzar la continuidad de estos esfuerzos durante la siguiente administración.

Asimismo, es importante hacer referencia a "IA2030Mx", una coalición multisectorial integrada por profesionales, instituciones académicas, empresas, organismos públicos y

[270] Véase Universidad Veracruzana, s.f..
[271] Véase Gobierno de Mexico, 2013.
[272] Véase Gobierno de Mexico, 2018.
[273] Véase Oxford Insights & C-Minds, 2018.
[274] Véase Gobierno de Mexico, 2018.

otros actores clave del ecosistema digital y de IA en México que fomenta el uso y la aplicación de tal tecnología en beneficio de los mexicanos; trata de fortalecer la coordinación y las sinergias entre distintos sectores; propicia debate sobre las oportunidades y los retos presentes y futuros relacionados con la IA y traduce tal debate en acciones prácticas; intenta aprovechar la trayectoria, el talento, la energía y el potencial de México para convertirlo en un país más competitivo y justo; e impulsa un conocimiento de la IA accesible a todos los ciudadanos. [275]

En Colombia, el 8 de noviembre del 2019 el Consejo Nacional de Política Económica y Social (CONPES), un organismo asesor del gobierno colombiano integrado en el Departamento Nacional de Planeación, publicó el denominado Documento CONPES, que contiene la "Política Nacional para la Transformación Digital e IA", un plan de política nacional para la transformación digital e implementación de la IA con el objetivo de potenciar el valor social y económico en el país a través del uso estratégico de tecnologías digitales, tanto en el sector público como en el sector privado, impulsar la productividad y favorecer el bienestar de los ciudadanos, de modo que Colombia pueda aprovechar las oportunidades y enfrentar los retos relacionados con la 4ª Revolución Industrial. Para alcanzar el objetivo establecido se prevé la disminución de las barreras que dificultan la implementación de las tecnologías digitales, tanto en el sector público como el privado; la creación de las condiciones adecuadas y necesarias para la innovación digital, como medio de aumento del valor económico y social; el fortalecimiento de las competencias del capital humano que faciliten la inserción de la sociedad colombiana en la 4ª Revolución Industrial; y el desarrollo de condiciones necesarias para preparar a Colombia para los cambios económicos y sociales que conlleva la IA, atribuyendo a esta tecnología un papel clave en la transformación digital.[276]

En Argentina, en julio de 2019, el gobierno anunció la preparación de una Estrategia Nacional de IA como parte de la Agenda Digital 2030, un plan de acción publicado en el Boletín Oficial el 5 de noviembre del 2018 con el foco puesto en la digitalización y en una implementación de la tecnología orientada hacia el ciudadano, la economía digital y los

[275] Gobierno de Mexico, s.f..
[276] Gobierno de Colombia, 2019.

empleos del futuro[277]. Asimismo, se publicó la iniciativa Argentina Innovadora 2030 y su Plan Nacional de Ciencia, Tecnología e Innovación (PNCTI).[278]

Las prioridades temáticas para la Estrategia Nacional de IA, cuya finalidad es fomentar el potencial económico de Argentina mediante la generación de condiciones favorables para el desarrollo y la implementación de tal tecnología en diferentes sectores de la industria y en varios niveles del gobierno, así como promover el desarrollo de una IA inclusiva y sostenible para una mejor calidad de vida de las personas, incluyen: talento y educación, I+D e innovación, infraestructuras de supercomputación y acciones para facilitar las transiciones laborales y la cooperación público-privada en el uso de datos, inversión, ética y regulación, comunicación, sensibilización y cooperación internacional, con previsión de desarrollo de un Centro Nacional de Innovación de IA para implementar proyectos [279].

En Brasil, el 21 de marzo del 2018 el Presidente del Gobierno firmó el denominado Decreto "E-Digital", que contenía la Estrategia Nacional de Digitalización brasileña, y disopnía un amplio conjunto de pautas y objetivos a largo plazo, especialmente en materia de IA.

Además, el Ministerio de Ciencia, Tecnología, Innovación y Comunicación durante el periodo 2014-2019 aportó incentivos y apoyo financiero para dieciséis proyectos sobre IA, casi sesenta nuevas empresas de tal ámbito, y cuarenta iniciativas para potenciar el uso de la IA en el gobierno federal. [280]

Finalmente, el 12 de abril de 2021 el Gobierno de tal país publicó su Estrategia Nacional de IA[281], que advierte de que Brasil se enfrenta a desafíos considerables para desarrollar su industria de IA, tales como la falta de mano de obra con conocimientos específicos, una farragosa burocracia y una elevada carga fiscal; se adhiere a los Principios de IA de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), que abogan por una IA centrada en el ser humano; y fija objetivos tales como: la formación de mano de

[277] Gobierno de Argentina, 2018.
[278] Véase Gobierno de Argentina, 2018.
[279] OECD, 2019.
[280] OECD, 2019.
[281] Véase Gobierno de Brasil, 2021.

obra cualificada; el impulso de la investigación, el desarrollo, la innovación y el espíritu empresarial; y la aplicación gubernamental de la IA.

En Chile, en agosto del 2019, la Comisión de Desafíos del Futuro, Ciencia, Tecnología e Innovación del Senado publicó el informe "*Hacia una Estrategia de IA para Chile*", que estableció los fundamentos y los motivos por los que Chile requiere una estrategia en IA, para lo cual, "*se describe la realidad nacional e internacional al respecto, se sintetizan estrategias de IA de algunos países, se destaca la posición de diferentes organismos internacionales en torno a la IA, se plasman los desafíos del país en términos de I+D y formación de personas, y su impacto en diferentes áreas, y se proponen los pasos a seguir para elaborar una estrategia de IA nacional*". [282]

Posteriormente, el 27 de agosto de 2019 el Presidente de la República encargó al Ministerio de Ciencia, Tecnología, Conocimiento e Innovación la elaboración de una Política Nacional de IA con la idea de que esta contuviera la estrategia que el país debería seguir en tal materia durante los siguientes diez años,[283] habiendo sido finalmente publicada el 28 de octubre de 2021.

-A NIVEL SUPRAESTATAL

En el ámbito internacional, y en concreto, intergubernamental, son también muchas las iniciativas que han ido surgiendo (y seguirán haciéndolo), de entre las cuales destacan las siguientes.

En mayo del 2018, el Comité de Política de Economía Digital de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), creó el "*Expert Group on AI in Society*"[284], integrado por más de cincuenta expertos de distintos sectores y disciplinas, con el fin de establecer los principios que deberían regir en las políticas públicas y la cooperación internacional, fundamentales para garantizar los derechos de los ciudadanos y fomentar la confianza en el ámbito de aplicación de la IA.

[282] Gobierno de Chile, 2019, pág. 2.
[283] Véase Gobierno de Chile, 2019.
[284] Véase OECD, 2019.

En mayo del 2019, la OCDE, basándose en los anteriores principios, adoptó "*the OECD Council Recommendation on AI*"[285], que contenía los denominados "*the OECD AI Principles*". En relación con ello procede advertir que, si bien las Recomendaciones de la OCDE no son jurídicamente vinculantes, estas son muy influyentes, y en este caso, tal y como veremos, han servido de base para la adopción de declaraciones de principios posteriores. En concreto, se establecen por la mencionada organización cinco principios complementarios, basados en valores y necesarios para llevar a cabo una implementación responsable de la IA de un modo confiable:

"*-la IA debería beneficiar a las personas y al planeta impulsando el crecimiento inclusivo, el desarrollo sostenible y el bienestar;*

-los sistemas de IA deben diseñarse de manera que respeten el Estado de Derecho, los derechos humanos, los valores democráticos y la diversidad, y deben incluir las garantías necesarias, por ejemplo, permitiendo la intervención humana cuando sea necesario, para asegurar una sociedad justa y equitativa;

-debe haber transparencia y hacerse una divulgación responsable de los sistemas de IA para garantizar que las personas entiendan los resultados basados en tal tecnología y puedan afrontarlos;

-los sistemas de IA deben funcionar de manera robusta y segura, y los riesgos potenciales deben evaluarse y gestionarse de forma contínua;

-las organizaciones y las personas que desarrollan, implementan u operan con sistemas de IA deben ser responsables de su correcto funcionamiento, de acuerdo con los principios anteriores."[286]

Y, asimismo, se disponen cinco recomendaciones para los reguladores y los responsables políticos:

-invertir en investigación y desarrollo de IA;
-fomentar un ecosistema digital para la IA;

[285] Véase OECD, 2019.
[286] OECD, 2019.

-formar un entorno político propicio para la IA;
-desarrollar la capacidad humana y prepararse para la transformación del mercado laboral;
-
cooperar internacionalmente para lograr una IA confiable.

Asimismo, se creó el OCDE AI Policy Observatory, un organismo valiosísimo que se ha convertido en un referente en el ámbito de la IA, puesto que centraliza y actualiza datos, y realiza análisis multidisciplinares basados en evidencias sobre aquellas áreas en que tal tecnología tiene un mayor impacto, habiéndose erigido como una fuente única de información y diálogo, como plataforma inclusiva para el diseño y la implementación de políticas públicas sobre IA. [287]

En el G20, el "*G20 Ministerial Meeting on Trade and Digital Economy*" celebrado en Tsubuka, Japón, los días 8 y 9 de junio del 2019, al que asistieron todos los miembros del G20 así como invitados tales como Países Bajos, España, Estonia, Chile, como economía anfitriona del Foro de Cooperación Económica Asia-Pacífico 2019, Egipto (en nombre de la Unión Africana), Nigeria, Senegal (en nombre de la Nueva Asociación para el Desarrollo), Singapur, Vietnam y algunas organizaciones internacionales, se aprobaron los "*G20 AI Principles*", publicados en un documento titulado "*G20 Ministerial Statement on Trade and Digital Economy*"[288], un compromiso del G20 de realizar una aproximación a la IA enfocada en el ser humano basado en los principios adoptados por los treinta y seis miembros de la OECD y seis países adicionales en mayo del 2019 [289] [290]. Entre otros, se hace referencia a los principios de seguridad y responsabilidad, transparencia, explicabilidad y robustez y, asimismo, se pone de manifiesto la necesidad de crear políticas nacionales y fomentar la cooperación internacional para generar una IA de confianza.

En el G7, en la "*G7 Leaders Summit: Digital Economy and Artificial Intelligence*" celebrada en Biarritz (Francia) del 24 al 26 de agosto del 2019, el Secretario General de la OCDE expuso:

[287] OECD, s.f..
[288] Véase G20, 2019.
[289] Véase OECD, 2019.
[290] Es importante resaltar que China y Rusia fueron parte de los participantes del G20 pero no fueron signatarios de los principios de la OCDE.

"*Con la adopción de los Principios de IA de la OECD, que se convirtieron en los Principios de IA del G20 en Osaka, hemos acordado colectivamente una política global y un punto de referencia ético que promueve la innovación, el crecimiento inclusivo, el desarrollo sostenible y los derechos humanos. Cada principio del G20 sobre IA es una hoja de ruta. Ahora, debemos centrarnos en la implementación, la implementación, y la implementación: el G7 debe liderar con el ejemplo y allanar el camino a seguir. No estamos con las manos vacías. Además de los principios de IA, tenemos:*

.el "OECD's AI Policy Observatory", un centro para la medición multidisciplinaria, basada en evidencia, análisis de políticas y diálogo, que nos ayudará a traducir este enfoque común en políticas públicas concretas;

.y la creación de una "*Global Partnership on AI" encabezada por la Presidencia francesa y Canadá, que informará los debates sobre políticas con una reflexión prospectiva, de alto nivel, y a largo plazo por parte de la comunidad científica, los expertos, académicos, institutos de investigación y la industria*. (…)"[291]

El 26 de agosto de ese mismo año, en el mismo contexto, los líderes del G7, junto con el Secretario General de la OCDE, firmaron una declaración de compromiso para avanzar de forma conjunta y determinada en la potenciación del buen uso de las nuevas tecnologías y la minimización de sus potenciales riesgos bajo el título "*Biarritz Strategy for an Open, Free and Secure Digital Transformation*"[292]. En concreto, respecto de la IA se dispuso: "*7. Las tecnologías de IA están provocando una transformación radical de nuestras sociedades y economías. Pueden abrir un ciclo de innovación y crecimiento sin precedentes. La IA puede proporcionar soluciones innovadoras para avanzar en el progreso hacia el logro de la Agenda 2030 para el Desarrollo Sostenible, así como beneficios significativos para ayudar a abordar algunos de nuestros desafíos más apremiantes. Los líderes reconocen que la IA está transformando las sociedades, la economía global y el futuro del trabajo y tiene el potencial de mejorar el bienestar de las personas, pero puede tener efectos dispares con respecto a la economía y la privacidad y la protección de datos, e implicaciones para la democracia.*"

[291] OECD, 2019.
[292] G7, 2019, pág. 2.

En la Organización de las Naciones Unidas (ONU), The International Telecommunication Union trabajó con otras más de veinticinco agencias de la ONU para organizar la Cumbre Global "*AI for Good Global Summit*" que, tras su primera edición en junio del 2017, se celebra de forma anual, habiéndose asociado con organizaciones como The XPRIZE Foundation y The Association for Computing Machinery.[293]

En septiembre de 2017, The United Nations Interregional Crime and Justice Research Institute (UNICRI) firmó "*The Host Country Agreement*" para abrir un Centro de IA y Robótica en La Haya (Países Bajos) con el objetivo contribuir a mejorar la comprensión de la dualidad riesgo-beneficio de la IA y la robótica a través de una mejor coordinación, recopilación y difusión de conocimientos, capacitación, y actividades de sensibilización y divulgación.[294]

En el mismo mes y el mismo año, el World Commission on Ethics of Scientific Knowledge and Technology (COMEST), perteneciente a la UNESCO, publicó "*the Report of the World Commission on the Ethics of Scientific Knowledge and Technology on Robotics Ethics*"[295], que tras establecer una definición de robot y su relación con los algoritmos y la IA, analizó el posible impacto de ello en diversos ámbitos de la sociedad (industria, defensa, sociedad civil, transporte, sanidad, educación, hogar, agricultura y medio ambiente), y poner de manifiesto los desafíos éticos que presentaBA, incluyendo una lista de "Principios y valores éticos relevantes", a saber: dignidad humana, valor de la autonomía, valor de la privacidad, principio de "no causar daño", principio de responsabilidad, valor de la beneficencia y valor de la justicia; y estableció una serie de recomendaciones específicas sobre ética en materia de robótica.

En septiembre del 2018 la UNESCO organizó una mesa redonda pública con expertos, y en marzo del 2019 celebró la conferencia: "*Principles for AI: Towards a Humanistic Approach? – AI with Human Values for Sustainable Development*", con el objetivo crear conciencia y promover la reflexión y el debate sobre las oportunidades y los desafíos que plantea la IA y las tecnologías relacionadas con ella.

[293] OECD, 2019.
[294] UNICRI, 2017.
[295] UNESCO, 2017.

En noviembre de 2019, en la 40ª Conferencia General de la UNESCO se anunció el desarrollo de una Recomendación sobre Ética en el ámbito de la IA que, tras haber sido debatida en el "*Intergovernmental Meeting related to the draft Recommendation on the Ethics of Artificial Intelligence*" celebrada el 21 de julio de 2021, se aprobó en la 41ª Asamblea General culminada el 24 de noviembre de 2021.

Y es que tal Recomendación sobre la Ética de la IA puede considerarse, sin duda, como un documento histórico, habida cuenta de que fue firmado por los ciento noventa y tres países que forman parte de la UNESCO (entre ellos, China, uno de los más díscolos en el panorama ético-tecnológico mundial), y constituye una guía ética con principios y valores comunes a tener en cuenta en la aplicación de la IA, lo cual supone un buen punto de partida.

Para ello, tal documento, tal y como afirma Audrey Azoulay, Directora General de la UNESCO, se basa "*en tres pilares: el respeto de los derechos humanos, el Estado de Derecho y la lucha contra la discriminación*"[296] y establece directrices para la regulación de técnicas de IA tan controvertidas como el reconocimiento facial.

No obstante, lamentablemente, tal texto no tiene carácter vinculante, si bien la UNESCO realizará controles y evaluaciones periódicas de su aplicación, requiriendo a los Estados miembros que informen sobre sus práctias y progresos, sometiendo los resultados, analizados por expertos, a debate público de forma transparente.

Asimismo, The Centre for Policy Research de la Universidad de las Naciones Unidas, tras el mandato conferido a esta por la "*Secretary-General's Strategy on New Technologies*", ha creado el *think tank* denominado "*AI and Global Governance Platform*"[297], un espacio inclusivo para investigadores, políticos, líderes corporativos y de pensamiento para explorar este desafío de política pública para analizar y generar soluciones innovadoras ante los desafíos actuales y futuros de las políticas públicas mundiales relativas a tal tecnología.

[296] Hidalgo, 2021.
[297] United Nations University, 2014.

En mayo de 2018, los ministros de los países nórdicos y bálticos (Dinamarca, Estonia, Finlandia, Islas Feroe, Islandia, Letonia, Lituania, Noruega, Suecia e Islas Åland) firmaron una declaración conjunta "*AI in the Nordic-Baltic Region*"[298] para reforzar su cooperación en el ámbito de la IA y reivindicar su posición como región líder de Europa en el área de desarrollo digital. La declaración hace referencia a siete áreas de trabajo para desarrollar y promover un uso de la IA beneficioso para los ciudadanos:

-la mejora de las oportunidades en el desarrollo de habilidades para que más autoridades, empresas y organizaciones utilicen la IA;
-la mejora del acceso a los datos con el fin de que la IA sea empleada para prestar mejores servicios a los ciudadanos y a las empresas de la región;
-el desarrollo de pautas, estándares, principios y valores éticos y transparentes que sirvan de guía en el cuándo y el cómo deben usarse las aplicaciones de IA;
-el intento de que la infraestructura, el *hardware*, el *software* y los datos, fundamentales para el uso de la IA se basen en estándares que permitan la interoperabilidad, la privacidad, la seguridad, la confianza, el buen uso y la portabilidad;
-la garantía de que la IA obtenga un lugar destacado en la discusión europea y la implementación de iniciativas en el marco del Mercado Único Digital;
-la evitación de regulaciones innecesarias; y
-la utilización del Consejo Nórdico de Ministros para facilitar la colaboración en áreas políticas relevantes.[299]

Ya en 1987 la Organización Internacional de Normalización (ISO) y la Comisión Electrotécnica Internacional (IEC) crearon un comité técnico conjunto ISO/IEC JTC 1, al que encomendaron el desarrollo de estándares de tecnología de la información para aplicaciones comerciales y de consumo. Posteriormente, en octubre de 2017, se creó, dentro del JTC 1, el subcomité 42 (SC 42)[300], para desarrollar estándares en el ámbito de la IA y proporcionar orientación a los comités ISO e IEC que desarrollan aplicaciones de tal tecnología.[301]

[298] Gobierno de Suecia & Consejo Nórdico de Ministros, 2018.
[299] OECD, 2019.
[300] Véase Price, 2018.
[301] OECD, 2019.

En el ámbito de la Unión Europea (UE), ya en 2013 se dictó el Reglamento nº 1291/2013, del Parlamento Europeo y del Consejo, que estableció el programa "Horizonte 2020", Programa Marco de Investigación e Innovación (2014-2020) que tenía como objetivo lograr en tal periodo un mayor impacto en la investigación y la innovación combinando los fondos de tal programa y del sector privado, dentro de asociaciones público-privadas, con el fin de alcanzar los objetivos generales de competitividad de la UE y ayudar a afrontar los retos de la sociedad en aquellos casos que presentaban un claro valor añadido europeo. En dicho Reglamento se establecía que la gobernanza y el funcionamiento del programa "Horizonte 2020" debían ser abiertos, transparentes, efectivos, eficientes e inclusivos, dando la posibilidad de participar a una amplia gama de partes activas interesadas en sus ámbitos específicos.

Asimismo, la IA fue incluida dentro de la Estrategia lanzada por la Comisión Europea para la digitalización de la industria el 19 de abril del 2016, a través de la Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones bajo el título: "Digitalización de la industria europea. Aprovechar todas las ventajas de un mercado único digital", así como en la posterior Resolución emitida por el Parlamento Europeo el 1 de junio de 2017, sobre la digitalización de la industria europea, y la Comunicación de la Comisión lanzada el 13 de septiembre del 2017 titulada "Invertir en una industria inteligente, innovadora y sostenible. Estrategia renovada de política industrial de la UE".

El Parlamento Europeo, en su resolución de 16 de febrero del 2017, con recomendaciones para la Comisión sobre obre normas de Derecho civil sobre robótica, a pesar de que puso de manifiesto que "*el potencial para el empoderamiento mediante el uso de la robótica está matizado por un conjunto de tensiones o riesgos y debe evaluarse seriamente desde el punto de vista de la seguridad y la salud; la libertad, privacidad, integridad y dignidad; la autodeterminación y no discriminación, y la protección de datos personales*", también subrayó la importancia del principio de transparencia y advirtió de que el marco ético rector debe basarse en los *"principios de buena fe, autonomía y justicia, en relación a los principios y valores consagrados en el artículo 2 del Tratado de la Unión Europea y en la Carta de los Derechos Fundamentales".*

El 25 de abril del 2018 la Comisión Europea dictó la Comunicación al Parlamento Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones, con el título "IA para Europa", bajo la creencia de que la UE "*debe adoptar un planteamiento coordinado que le permita aprovechar al máximo las oportunidades que brinda la IA y abordar los nuevos retos que conlleva. La UE tiene la posibilidad de liderar el desarrollo y la utilización de la IA de una vez y para todos, partiendo de los valores y puntos fuertes con los que cuenta*".[302] Tal estrategia de IA para Europa tiene como fin:

-"*Potenciar la capacidad tecnológica e industrial de la UE e impulsar la adopción de la IA en todos los ámbitos de la economía, tanto en el sector privado como en el público* (…)*;*
-prepararse para las transformaciones socioeconómicas que origina la IA, fomentando la modernización de los sistemas de educación y formación, favoreciendo el talento, previendo los cambios en el mercado laboral y prestando apoyo a las transiciones que se operen en él y a la adaptación de los sistemas de protección social; y
-garantizar el establecimiento de un marco ético y jurídico apropiado, basado en los valores de la Unión y en consonancia con la Carta de los Derechos Fundamentales de la UE. Incluye una próxima directriz sobre la interpretación de las actuales normas en materia de responsabilidad por productos defectuosos y un análisis pormenorizado de los retos emergentes, así como la cooperación con las partes interesadas, en el seno de una Alianza europea de la IA, para elaborar directrices éticas en la materia."[303]

En virtud de tal estrategia, el 7 de diciembre del mismo año, la Comisión Europea presentó la Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones titulada: "Plan coordinado sobre la IA", un plan conjunto preparado junto con los Estados miembros para fomentar el desarrollo y la utilización de la IA en Europa que "*propone actuaciones conjuntas para lograr una cooperación más estrecha y eficiente entre los Estados miembros, Noruega, Suiza y la Comisión en cuatro ámbitos clave: aumentar la inversión, lograr que haya más datos disponibles, fomentar el talento y garantizar la confianza*." y establece que "*Es fundamental reforzar la coordinación para que Europa se convierta en*

[302] Pág. 2.
[303] Pág. 4.

la región que dirija a nivel mundial la creación e implantación de una inteligencia artificial puntera, ética y segura."[304]

El 6 de junio del 2018 se publicó la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecía el programa "Europa Digital para el periodo 2021-2027", que estableció como objetivos:

"*-desarrollar y fortalecer las capacidades de informática de alto rendimiento y de procesamiento de datos de la UE y garantizar su amplio uso tanto en áreas de interés público como la salud, el medio ambiente y la seguridad, como por la industria, en particular las pymes;*
-desarrollar y reforzar las capacidades esenciales IA como los recursos de datos y las bibliotecas de algoritmos de inteligencia artificial y hacerlos accesibles a todas las empresas y administraciones públicas, así como reforzar y fomentar los vínculos entre las instalaciones de ensayo y experimentación en inteligencia artificial existentes en los Estados miembros;
-velar por que las capacidades esenciales necesarias para la seguridad de la economía digital, la sociedad y la democracia de la UE estén presentes y sean accesibles para el sector público y las empresas de la UE, y mejoren la competitividad de la industria de la ciberseguridad de la UE;
-velar por que la población activa actual y futura pueda adquirir fácilmente competencias digitales avanzadas, especialmente en informática de alto rendimiento, inteligencia artificial y ciberseguridad, ofreciendo a estudiantes, titulados y personal en activo, independientemente de donde se encuentren, los medios para obtener y desarrollar dichas competencias;
-extender el mejor uso de las capacidades digitales, especialmente la informática de alto rendimiento, la inteligencia artificial y la ciberseguridad, al conjunto de la economía, en áreas de interés público y la sociedad, incluido el despliegue de soluciones interoperables en áreas de interés público y facilitar el acceso a la tecnología y al conocimiento a todas las empresas, en particular a las pymes."

[304] Comisión Europea, 2018.

El 11 de septiembre del 2018 el Parlamento Europeo dictó la Resolución sobre la igualdad lingüística en la era digital habiendo dispuesto "*que el desarrollo de las tecnologías del lenguaje abarca numerosos ámbitos y disciplinas de investigación, incluidas la lingüística computacional, la IA, la informática y la lingüística (con aplicaciones como el tratamiento del lenguaje natural, el análisis de texto, la tecnología de voz y la minería de datos, entre otras)*".[305]

El 28 de septiembre del 2018 el Consejo de la Unión Europea dictó el Reglamento (UE) 2018/1488, por el que se creó la Empresa Común de Informática de Alto Rendimiento Europea, con el fin de dotar a la UE del rendimiento informático necesario para que su investigación se mantuviera a la vanguardia, con coordinación de la inversión de los Estados miembros en informática de alto rendimiento y el refuerzo de la incorporación de tal tecnología por parte de la industria y el mercado tanto en el sector público como en el privado.

Asimismo, el 12 de septiembre del 2018 el Parlamento Europeo aprobó la Resolución sobre los sistemas armamentísticos autónomos, entendida necesaria bajo la consideración de que "*por «sistemas armamentísticos autónomos letales» se entienden sistemas de armas sin un control humano significativo con respecto a las funciones críticas de selección y ataque de objetivos individuales*"[306] y que "*las políticas y acciones de la Unión se inspiran en los principios de los derechos humanos y el respeto de la dignidad humana, en los principios de la Carta de las Naciones Unidas y en el Derecho internacional; que esos principios han de aplicarse a fin de preservar la paz, prevenir los conflictos y reforzar la seguridad internacional*".[307]

El 12 de febrero del 2019 el Parlamento Europeo dictó la Resolución sobre una política industrial global europea en materia de IA y robótica, bajo la consideración de que "*una IA y una robótica transparentes y que integren consideraciones éticas tienen el potencial necesario para enriquecer nuestras vidas y consolidar nuestras capacidades, tanto en el plano individual como para el bien común*"[308] y que "*la integración cada vez mayor de la*

[305] Considerando B.
[306] Considerando B.
[307] Considerando A.
[308] Considerando A.

robótica en los sistemas humanos requiere una fuerte orientación normativa sobre el modo de maximizar las ventajas y de reducir los riesgos para la sociedad, así como de garantizar un desarrollo seguro y equitativo de la IA".[309]

En marzo del 2019 el Parlamento Europeo publicó el informe "*Understanding algorithmic decision-making: Opportunities*" [310], que analizaba las oportunidades y los riesgos relacionados con el uso de sistemas de toma de decisiones mediante algoritmos, y proponía políticas para reducir los riesgos, explicar bien sus limitaciones y poder beneficiarse de las enormes y prometedoras posibilidades que ofrecen estos sistemas.

Posteriormente, el 8 de abril del 2019 la Comisión Europea publicó la Comunicación al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones, bajo el título "Generar confianza en la IA centrada en el ser humano". Las directrices que se propugnaron para lograr una "IA fiable", fueron fundamentalmente tres:

- ✓ cumplimiento de la ley,
- ✓ respeto de los principios éticos,
- ✓ y solidez.

Además, se dispuso que era necesario, en todo caso, comunicar a los usuarios que estaban interactuando con un sistema de IA y garantizar la trazabilidad del mismo, debiendo registrar y documentar la totalidad del proceso que hubiera dado lugar a la toma de decisiones, con inclusión de una descripción del modo de obtención de los datos empleados y su etiquetado, así como una explicación del algoritmo empleado en cada caso y el grado en que había influido en la toma de decisiones, de forma comprensible para las personas afectadas, con el fin de que lo pudieran entender sin que fueran necesarios conocimientos técnicos.

El 19 de febrero del 2020 la Comisión Europea publicó el denominado "Libro Blanco sobre IA. Un enfoque europeo orientado a la excelencia y la confianza", que puso el foco en el rápido desarrollo de la IA y en su poder para cambiar nuestras vidas, tanto con posibles

[309] Considerando C.
[310] Véase Parlamento Europeo, 2019.

beneficios (mejora de la atención sanitaria, aumento de eficiencia en la agricultura, ayuda a la mitigación del cambio climático, e incremento de la seguridad de los europeos, entre otros) y sus potenciales riesgos, como en la opacidad en la toma de decisiones, la discriminación de género o de otro tipo, la intromisión en nuestras vidas privadas, o su uso con fines delictivos.

Por su parte, en el mes de julio de 2020, se creó en el Parlamento Europeo el denominado Comité de IA (del que forma parte la eurodiputada española Pilar del Castillo Vera), con la finalidad de analizar el impacto y los desafíos de implementar tal tecnología en el ámbito de la UE[311]. Entre otras iniciativas, el Presidente de dicho Comité, el eurodiputado rumano Dragoș Tudorache aboga por la creación de un "registro de riesgos" de las distintas aplicaciones de IA.

En relación con ello, y un contexto de feroz competencia mundial, en el mencionado Libro Blanco se recoge la necesidad de que la UE establezca un enfoque sólido basado en la Estrategia Europea para la IA presentada en abril de 2018, con el fin de aprovechar las antedichas oportunidades que tal tecnología ofrece y abordar los retos que presenta, con una actuación conjunta y la determinación de la forma en que, a partir de los valores europeos, promoverá su desarrollo y adopción.

Y, finalmente, el 21 de abril de 2021 la Comisión Europea, en el marco del proyecto de la UE de regular la IA de forma específica y convertirse en el centro mundial de una IA ética y fiable, publicó su propuesta al Parlamento Europeo y al Consejo de nuevas medidas y normas bajo el título Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión

Dichas normas, basadas en el nivel de riesgo que pueden entrañar los distintos sistemas de IA, establece una especial regulación para cada uno de ellos. Así, se distingue entre los sistemas de:

[311] Véase Parlamento Europeo, 2020.

-riesgo inadmisible (Título II, artículo 5), es decir, que impliquen una amenaza clara para los derechos fundamentales, la seguridad o los medios de subsistencia de los ciudadanos (entre otros, aquellos sistemas que posibiliten a los gobiernos la "puntuación social" de sus nacionales), que deberán resultar prohibidos;

-alto riesgo (Título III, Capítulo 1, artículos 6 y 7), es decir, que "*abarcan las tecnologías de IA empleadas en:*

.infraestructuras críticas (por ejemplo, transportes), que pueden poner en peligro la vida y la salud de los ciudadanos;
.formación educativa o profesional, que pueden determinar el acceso a la educación y la carrera profesional de una persona (por ejemplo, puntuación en exámenes);
.componentes de seguridad de los productos (por ejemplo, aplicación de IA en cirugía asistida por robots);
.empleo, gestión de trabajadores y acceso al trabajo por cuenta propia (por ejemplo, programas informáticos de clasificación de CV para procedimientos de contratación);
.servicios públicos y privados esenciales (por ejemplo, sistemas de calificación crediticia que priven a los ciudadanos de la oportunidad de obtener un prestamo);
.aplicación de las leyes, que pueden interferir con los derechos fundamentales de las personas (por ejemplo, evaluación de la fiabilidad de las pruebas);
.gestión de la migración, el asilo y el control de las fronteras (por ejemplo, comprobación de la autenticidad de los documentos de viaje);
.Administración de Justicia y procesos democráticos (por ejemplo, aplicación de la ley a un conjunto concreto de hechos)."[312]

Tales sistemas, según la Propuesta de la Comisión, deberán estar sometidos a obligaciones y controles estrictos antes de su comercialización (Título III, Capítulo 2, artículos 8 a 15, Capítulo 3, artículos 16 a 29; Capítulo 4, artículos 30 a 39; y Capítulo 5, artículos 40 a 51), a saber:

"*.sistemas adecuados de evaluación y mitigación de riesgos;*

[312] Comisión Europea, 2021.

.alta calidad de los conjuntos de datos que alimentan el sistema para minimizar los riesgos y los resultados discriminatorios;
.registro de la actividad para garantizar la trazabilidad de los resultados;
.documentación detallada que aporte toda la información necesaria sobre el sistema y su finalidad para que las autoridades evalúen su conformidad;
.información clara y adecuada al usuario;
.medidas apropiadas de supervisión humana para minimizar el riesgo;
.alto nivel de solidez, seguridad y precisión.";[313]

-riesgo limitado (Título IV, artículo 52), que deberán estar sometidos a específicas obligaciones de transparencia (entre otros, por ejemplo, los *chatbots*, siendo que los usuarios deberán ser informados de que están interactuando con una máquina para que puedan decidir libremente si desean continuar o no);

-riesgo mínimo o nulo, es decir, que no representan amenaza alguna para los derechos fundamentales, la seguridad o los medios de subsistencia de los ciudadanos (entre otros, por ejemplo, los filtros de *spam* del correo electrónico), por lo que pueden ser comercializados y empleados de forma gratuita y libre.

Además de lo expuesto, la Comisión propone la creación de un "Comité Europeo de IA" (Título VI, artículos 56 a 58) que vele de forma centralizada por la aplicación de tales normas e impulse la adopción de nuevas medidas en materia de IA, lo cual bajo mi punto de vista es una buenísima noticia, habida cuenta de que la existencia de un organismo así resulta fundamental para garantizar el cumplimiento de la normativa y asegurar el buen uso de los sistemas de IA. De hecho, tal propuesta se acerca a la idea de crear una especie de "Agencia Europea de IA" a la que ya he hecho referencia en páginas anteriores, si bien entiendo que para que la utilidad de dicho Comité sea real, estedebería asumir funciones y atribuciones no solo jurídicas sino también técnicas, así como contar con expertos multidisciplinares que permitieran establecer un control y un filtro real y completo a los sistemas de IA que quisieran operar y ser comercializados en el ámbito de la UE.

[313] Comisión Europea, 2021.

Asimismo, y de modo complementario, se sugiere un control de la aplicación de las nuevas normas y medidas por parte de las autoridades nacionales de los Estados Miembros, delegando la imposición de sanciones para casos de incumplimiento (Título VI, Capítulo 2, artículo 59 y Título X, artículo 71); se propone la creación de *sandboxes* (Título V, artículos 53 y 54) como espacios de innovación controlada y responsable; se sugiere la creación de unas guías o códigos de conducta (Título IX, artículo 69) de aplicación voluntaria para aquellos sistemas de IA que no impliquen alto riesgo; y se establece para las autoridades nacionales competentes y demás agentes implicados un deber general de confidencialidad sobre la información y los datos obtenidos en el desempeño de sus funciones (Título X, artículo 70).

Tras ello, el Parlamento Europeo y el Consejo son los que deben adoptar las propuestas de la Comisión Europea a través del procedimiento legislativo ordinario, lo que con toda probabilidad dará lugar a la publicación del tan esperado y necesario Reglamento de IA que será de aplicación directa en los Estados Miembros de la Unión y que, sin duda, sentará por fin las bases jurídicas del uso de la IA en la UE en el presente y en el futuro.

El Consejo de Europa ha sido (y continúa siendo), sin duda, uno de los organismos más comprometidos y activos en el ámbito de la IA.

Interesante es "*The Council of Europe Strategy for the rights of the child (2016-2021)*"[314], que incluye un análisis sobre los derechos del niño en Internet, estableciendo un plan de acción que tiene en cuenta que las nuevas tecnologías, entre ellas la IA, inevitablemente van a causar un impacto en el bienestar de los niños. Y, el Comité *ad hoc* para los derechos de los niños, en tal sentido, llevó a cabo la preparación de la "*Recommendation CM/Rec(2018)7 of the Committee of Ministers to member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment*".

En 2017, bajo la supervisión del Comité Directivo de la Sociedad de la Información y los Medios de Comunicación, el Comité de expertos en intermediarios de Internet (MSI-NET) elaboró un estudio sobre las dimensiones de los derechos humanos en las técnicas

[314] Consejo de Europa, 2016.

automatizadas de procesamiento de datos (en particular, algoritmos) y las posibles implicaciones regulatorias.[315] Como seguimiento, el Comité de expertos sobre las dimensiones de los derechos humanos del procesamiento automatizado de datos y las diferentes formas de IA está preparando un proyecto de recomendación sobre los impactos en los derechos humanos de los sistemas algorítmicos. El Comité, además, publicó en septiembre de 2019 un estudio sobre las implicaciones de la IA en el concepto de responsabilidad dentro de un marco de derechos humanos titulado "*A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework*".[316]

En la 526a Sesión Plenaria del Comité europeo Económico y Social, celebrada entre el 31 de mayo y el 1 de junio de 2017, se publicó un dictamen sobre el impacto social de la IA titulado "*The consequences of AI on the (digital) single market, production, consumption, employment and society*"[317], que instó a las partes interesadas de la UE a garantizar que el desarrollo, el despliegue y el uso de la IA estén orientados al bienestar social, con exigencia de estándares éticos, estrategias laborales adaptadas y una infraestructura europea de IA con entornos de aprendizaje de código abierto, habiéndose creado un grupo de estudio temporal sobre IA para examinar tales cuestiones.

En la Asamblea Parlamentaria celebrada el 28 de abril del 2017 se dictó la Recomendación 2102 bajo el título "*Technological convergence, Artificial Intelligence and Human Rights*"[318] , que instó al Comité de Ministros a que animara a los órganos pertinentes del Consejo a considerar cómo los artefactos inteligentes y/o dispositivos conectados y el *big data*, entre otros, desafiaban las diferentes dimensiones de los derechos humanos, proponiendo que se elaboraran directrices sobre las siguientes cuestiones:

-fortalecer la transparencia, la regulación por parte de las autoridades públicas y la responsabilidad de los operadores;
-establecer un marco común de normas que deban cumplirse cuando un tribunal utilice IA;

[315] Consejo de Europa, 2017.
[316] Véase Consejo de Europa, 2017.
[317] Véase Consejo de Europa, 2017.
[318] Véase Consejo de Europa, 2017.

-garantizar que cualquier máquina, robot o artefacto de IA permanezca bajo el control humano;
-reconocer nuevos derechos en relación con el respeto de la vida privada y familiar, el derecho a negarse a ser sometido a la elaboración de perfiles, a que se rastree la ubicación, a ser manipulado o influenciado por un *coach* y el derecho a tener la oportunidad, en el ámbito de atención y asistencia prestada a personas mayores o con discapacidad, para elegir tener contacto con un ser humano en lugar de un robot.[319]

En 2018 la Comisión Europea contra el Racismo y la Intolerancia publicó el estudio "Discriminación, IA y Toma de Decisiones Algorítmicas"[320], preparado por el Profesor Frederik Zuiderveen Borgesius para el Departamento de Antidiscriminación del Consejo de Europa, que hace referencia a los riesgos de discriminación causados por la toma de decisiones algorítmicas y otros tipos de IA, y fue la base de la declaración realizada por Christian Ahlund, miembro de la Oficina de ECRI, sobre IA y Democracia, en la "*High Level Conference on AI*" celebrada en Helsinki el 26 de febrero de 2019.

Asimismo, en octubre del mismo año, la Comisión de Igualdad de Género adoptó un proyecto de Recomendación al Comité de Ministros para prevenir y combatir el sexismo[321], que incluía pautas sobre cómo evitar los potenciales riesgos de que la tecnología perpetúe y aumente los prejuicios de género existentes y cómo conseguir aprovechar la IA para ayudar a cerrar brechas de género.

En su 93ª sesión Plenaria, celebrada entre los días 14 y 16 de novimbre, el Comité Europeo de Cooperación Jurídica, publicó un estudio técnico denominado "*Technical Study on Online Dispute Resolution Mechanisms*"[322] sobre la resolución de disputas en línea y el cumplimiento del derecho a un juicio justo y a un recurso efectivo, en relación con los artículos 6 y 13 del Convenio Europeo de Derechos Humanos (CEDH). Y el 16 de junio de 2021 publicó un documento que contenía directrices destinadas a garantizar la

[319] Págs. 1-2.
[320] Véase Consejo de Europa, 2018.
[321] Véase Consejo de Europa, 2018.
[322] Véase Consejo de Europa, 2018.

compatibilidad de los mecanismos de ODR con los artículos 6 y 13 del mencionado CEDH.[323]

El 13 de febrero del 2019 el Comité de Ministros aprobó la "*Declaration on the manipulative capabilities of algorithmic processes*"[324] bajo la conciencia del riesgo que implica el hecho de que las herramientas actuales de *Machine Learning* tengan la creciente capacidad no solo de predecir elecciones, sino también de influir en las emociones y los pensamientos y alterar decisiones humanas, a veces de forma subliminal, lo cual puede tener efectos significativos en la autonomía cognitiva de los individuos y en su derecho de formar opiniones y tomar decisiones independientes.

En febrero de 2019, asimismo, el Departamento de la Sociedad de la Información, en cooperación con el Gobierno finlandés, bajo la presidencia finlandesa del Comité de Ministros, organizó una conferencia de alto nivel "*AI: Governing the Game Changer. Impacts of AI Development on Human Rights, Democracy and the Rule of Law*", que reunió a expertos y responsables políticos de múltiples disciplinas para discutir los impactos del desarrollo de la IA en los derechos humanos, la democracia y el Estado de Derecho, y explorar opciones de la acción coordinada para garantizar que existan controles adecuados y una supervisión democrática.

En mayo del 2019, el Comisionado de Derechos Humanos del Consejo de Europa publicó una Recomendación titulada: "*Unboxing Artificial Intelligence: 10 steps to protect Human Rights*"[325], que proporcionó orientación sobre la forma en que se puede prevenir o mitigar el impacto negativo de los sistemas de IA en los derechos humanos centrándose en diez áreas clave de acción.

El 11 de septiembre de 2019, por su parte, el Comité de Ministros creó un Comité *ad hoc* sobre IA, el denominado CAHAI[326], con el objetivo de examinar la viabilidad y los potenciales riesgos de tal tecnología a través de amplias consultas con las partes

[323] Véase Consejo de Europa, 2018.
[324] Véase Consejo de Europa, 2019.
[325] Véase Consejo de Europa, 2019.
[326] Véase Consejo de Europa, 2019.

interesadas, y establecer un marco legal para el desarrollo, el diseño y la aplicación de la IA basado en los estándares del Consejo de Europa sobre derechos humanos, democracia y Estado de Derecho.

Además de todo lo anterior, es interesante poner el foco en uno de los proyectos más ambiciosos, necesarios e idóneos, bajo mi punto de vista, que tiene entre manos el Consejo de Europa: la creación de un mecanismo de certificación para productos de IA utilizados en sistemas judiciales, en relación con la Carta ética Europea sobre el uso de la IA en dichos sistemas anteriormente mencionada. Tal proyecto, presentado en la reunión del Comité Europeo para la Eficiencia de la Justicia (CEPEJ) celebrada en Atenas el día 23 de septiembre del 2019, está dirigido tanto al sector público como al sector privado, que se beneficiarán de asesoramiento metodológico y operativo sobre el modo de aplicar cada principio de la Carta.

En 2019, "*the Committee on Legal Affairs and Human Rights*" decidió crear un subcomité de IA y derechos humanos.

Finalmente, es importante poner de relieve que el Comité Europeo de Cooperación Jurídica (CDCJ) está trabajando actualmente en los mecanismos de resolución de disputas en línea y está tomando en consideración las posibles aplicaciones de la IA en dichos sistemas y su cumplimiento del derecho a un juicio justo. En relación con ello, en 2019 se estableció un grupo de redacción entre sus miembros para desarrollar proyectos que fijaran directrices para la atención de los responsables de políticas de diseño de mecanismos de resolución de disputas en línea con el fin de garantizar la compatibilidad de dichos mecanismos con los artículos 6 y 13 del CEDH.

2.4.2.3. ¿HACIA UNA REGULACIÓN GLOBAL DE LA IA?

A la vista de la prolija, extensa, creciente y diversa regulación sobre IA existente en el mundo (de la que se ha expuesto una muestra ciertamente representativa, si bien no completa, como consecuencia de su gran cantidad y de los constantes cambios que se producen en tal sentido) que puede, sin duda, resultar altamente beneficiosa, pero también altamente compleja y poco práctica, me planteo hacia dónde debe enfocarse el futuro del

marco jurídico aplicable en el ámbito de tal tecnología para que esta devenga lo más útil posible para la humanidad.

Y es que, tras el análisis de la regulación y de las iniciativas de distinta índole expuestas, he advertido la voraz carrera existente en la actualidad entre los distintos países del mundo por liderar la tranformacion digital y, en conreto, por estar a la cabeza de la implementación y el desarrollo de la IA en nuestras sociedades, habida cuenta de las enormes oportunidades que esta brinda, y que, sin duda, se traducen en poder.

Como consecuecia de ello, bajo mi punto de vista, la opción más idónea y acertada sería la de fijar, tal y como ya se está haciendo, unos principios básicos internacionales en materia de IA, sin perjuicio, no obstante, de su ulterior desarrollo a nivel nacional o incluso supranacional. Y es que los seres humanos debemos tomar conciencia colectiva y adoptar de forma conjunta decisiones (a poder ser preventivas, no reactivas) sobre cuestiones que nos atañen y nos afectan a todos, como especie, como ocurre en el ámbito medioambiental, por ejemplo, puesto que "la unión hace la fuerza" y, en caso contrario, los efectos negativos podrían ser globales e irreversibles.

Tal y como ya se ha expuesto en el punto 2.4.1, las nociones de ética y moral resultan muy distintas en cada cultura, por lo que puede reputarse utópica la idea de unificar criterios o establecer un marco jurídico y de actuación común de forma total. No obstante, tal y como se desprende de los distintos planes, iniciativas, estrategias y regulaciones examinadas, existe una premisa que prácticamente comparte toda la humanidad, y que, bajo mi punto de vista, es la que debe servir de punto de conexión para establecer unos principios de IA globales: la IA debe emplearse en beneficio de la humanidad, nunca en perjuicio de la misma.

Y es que no resulta verosímil pensar en un ser humano capaz de permitir la pérdida del control que actualmente ostenta sobre el mundo y sobre su propia especie (que, no obstante, no resulta completo, ni mucho menos, ya que es bien sabido que la naturaleza nos pone en nuestro sitio cuando menos lo esperamos). Así, con el fin de garantizar nuestra supervivencia y continuidad como especie y mantener el ya escaso control que ostentamos sobre la misma, entiendo que resulta necesaria la aprobación por parte de los máximos

países posibles (en el marco de la ONU o cualquier otra institución internacional, preexistente o creada *ad hoc*), de forma similar a lo que se hizo el 10 de diciembre de 1948 con la Declaración Universal de los Derechos Humanos, de una especie de "*Bill of Rights*" que detalle unos mínimos éticos y morales que los sistemas de IA deban cumplir y se erija como un verdadero escudo de protección frente a las nuevas criaturas tecnológicas que amenazan con sustituirnos y dejar a la especie humana vacía de contenido.

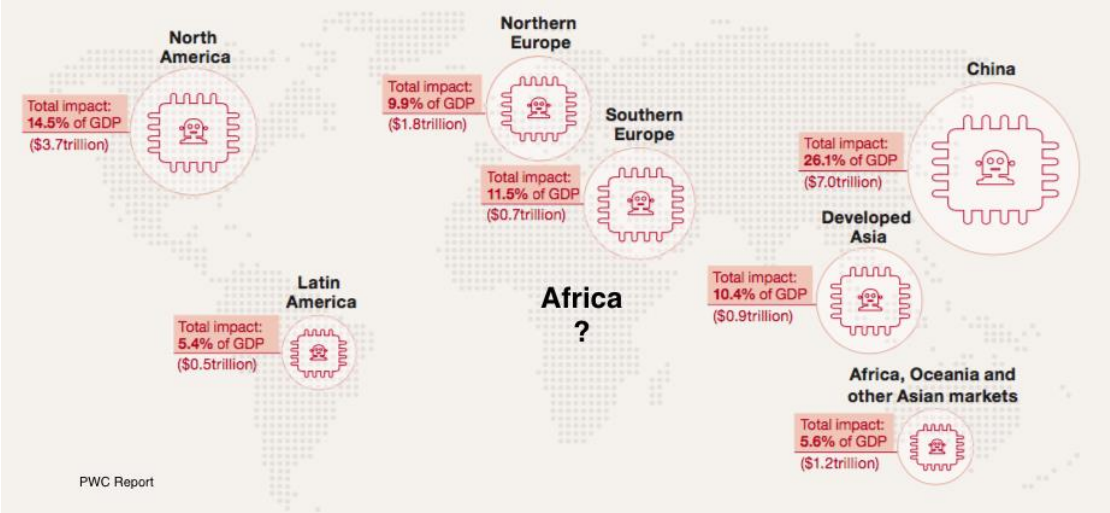
No obstante, tal y como ya se ha visto en el punto 2.4.2, han sido ya publicadas múltiples declaraciones de principios con tal objetivo, si bien bajo mi punto de vista, se requiere dar un paso más, y firme, puesto que, si no, existe el riesgo de que su valor resulte meramente abstracto y poco efectivo. Y al hablar de "paso firme" hago referencia a la necesidad de que la Declaración Universal de principios y valores de IA a la que se alude tenga un carácter jurídicamente vinculante para los países firmantes (que, tal y como se ha avanzado, deberían ser cuantos más posibles), estableciendo las correspondientes sanciones para casos de incumplimiento y fijando un sistema de prevención y disuasión infractora extremadamente potente, ya que hay que tener en cuenta que una vez transgredidas las bases internacionalmente acordadas podría resultar prácticamente imposible recuperar el control y depurar responsabilidades.

Es evidente que, como siempre, hay países muy reticentes a adoptar compromisos de tal calado internacional y, sobre todo, tan limitativos de soberanía y de poder como los que se proponen, pero en este caso es sorprendente la gran acogida que está teniendo la llamada internacional de consenso (otra cosa, no obstante, será el resultado).

Cuestión fundamental, tal y como he advertido, es que el proceso de preparación y creación de un marco jurídico común para el desarrollo y uso de la IA sea inclusivo y permita, facilite y fomente la participación del mayor número posible de países, especialmente del continente africano, que es el que, por el momento, está a la cola en el ámbito de tal tecnología y, sin embargo, tiene un potencial enorme. Y es que su exclusión podría implicar un aumento exponencial de la brecha de desigualdad ya existente con el resto del mundo (especialmente, como se expondrá más adelante, por las captaciones masivas de datos). Así, si bien la mayor parte de los centros y laboratorios de desarrollo y diseño de IA se concentran en ricas burbujas de innovación tales como Silicon Valley (EEUU) y

Zhongguancun (China)[327], la carrera conjunta hacia la consecución de una IA que garantice la soberanía del ser humano y respete los derechos fundamentales del mismo tiene que abrir fronteras, ya que ello no solo puede tener efectos muy positivos para el bienestar de unas pocas sociedades, en detrimento de otras, puesto que la consecución de ciertos objetivos globales solamente puede alcanzarse con la contribución de todos, como ocurre con el necesario control del cambio climático, que tanta amenaza supone para nuestro planeta.

Respecto de ello, en 2017, la consultora Price Waterhouse Coopers (PWC) emitió un estudio sobre el impacto económico de la IA en la economía mundial para 2030[328] que puso de manifiesto que tal tecnología puede llegar a aumentar el PIB global en un 14%, lo que la convierte en la mayor oportunidad comercial en la economía mundial. No obstante, la situación del continente africano marca una clara diferencia con la tendencia del resto de continentes del planeta.



329

En relación con la integración del continente africano en el ámbito de la IA, cierto es que cada vez van surgiendo más iniciativas, de naturaleza fundamentalmente privada, orientadas a ello. Así, a finales de 2013, IBM Research abrió su primera oficina africana

[327] Véase MIT Technology Review, 2019.
[328] Véase Price Waterhouse Coopers, 2017.
[329] Price Waterhouse Coopers, 2017.

en Nairobi (Kenia), y en 2016 procedió a la apertura de la segunda en Johannesburgo (Sudáfrica).

También en 2013, un grupo local de profesionales e investigadores lanzó el "*Data Science Africa*", un foro de debate y discusión creado con el objetivo de compartir y analizar recursos e ideas relacionadas con el *data science*, habiéndose celebrado el primer evento en la Universidad Tecnológica Dedan Kimathi de Nyeri (Kenia), en el año 2015.[330]

En 2017, otro grupo local formó la organización Deep Learning INDABA, una organización cuya misión es fortalecer y desarrollar el *Machine Learning* y la IA en África, con el fin de que los africanos dejen de ser meros observadores y receptores de los avances que van surgiendo en tales ámbitos en otros continentes, y se conviertan en activos diseñadores y propietarios de los mismos.[331]

En 2018, con la intención de avanzar en la creación de una industria de tecnologías de la información y comunicación (TIC) sólida y robusta, lo que constituye sin duda la base para un ulterior posicionamiento en el ámbito de la IA, en Kigali (Ruanda), The African Institute for Mathematical Sciences (AIMS) creó un programa de formación de un año "*African Master's in Machine Intelligence*" (AMMI) en asociación con Facebook y Google para formar a la próxima generación de líderes tecnológicos.[332]

En diciembre del mismo año, la UNESCO organizó su primera gran conferencia internacional sobre IA en Benguérir (Marruecos), con más de 400 participantes, incluidos expertos y representantes de alto nivel de los sectores público y privado, que examinaron formas de utilizar la IA para catapultar su desarrollo en África[333].

En abril del 2019, Google abrió su primer centro de investigación de IA en Ghana, con sede en Accra, con el objetivo de desarrollar soluciones que ayuden a mejorar la atención médica, la agricultura y la educación, si bien hoy en día todavía existen múltiples barreras

[330] Véase Data Science Africa, 2013.
[331] Véase Deep Learning INDABA, 2017.
[332] Véase The African Institute for Mathematical Sciences, 2018.
[333] Véase UNESCO, 2019.

que impiden que este funcione con normalidad. Y es que, entre otras, a las dificultades existentes para viajar por el interior del país, se unen los obstáculos para salir del continente africano o asistir a conferencias internacionales en caso de ser nacional de algún país del mismo, tal y como asegura su Director Moustapha Cisse, original de Senegal: "*A pesar del apoyo, muchos de nosotros todavía tenemos problemas para llegar a las conferencias. Me aceptaron en las reuniones pero no pude asistir porque países occidentales como Australia me negaron una visa, a pesar de que ya estaba establecido y trabajando profesionalmente en Europa*. (…) *Necesitamos más esfuerzos para superar estas barreras y garantizar que los beneficios de la IA lleguen a nivel mundial*".[334]

Y en junio del 2019 la cumbre líder mundial sobre derechos humanos en la era digital RightsCon, organizada de forma anual por la ONG AccessNow, fue celebrada en Túnez (Túnez), con casi tres mil participantes registrados, provinientes de más de ciento veinte países y más de setecientas cincuenta organizaciones asistentes.[335]

Para 2020 la International Conference on Learning Representations (ICLR), la principal reunión de profesionales dedicados a la rama de la IA denominda *Deep Learning*, estaba programada para celebrarse en abril en Addis-Abeba (Etiopía), a pesar de que la crisis sanitaria por el COVID-19 obligó a posponerla.

A pesar de las crecientes expectativas de difusión de la IA en África y de los esfuerzos realizados, no obstante, la falta de una infraestructura tecnológica adecuada y fácilmente disponible frena en seco el progreso en dicho continente.

En relación con lo expuesto con anterioridad, bajo mi punto de vista, la ONU sería sin duda el marco más idóneo para establecer una Declaración Universal de Principios sobre IA, habida cuenta de que es la organización internacional que cuenta con más miembros en sus filas (todos los países oficiales e independientes del mundo salvo la Ciudad del Vaticano, que ostenta rol de Estado observador). Y, en tal sentido, la UNESCO ya se ha postulado como centro de diálogo y debate para establecer las bases de la gobernanza global de la IA, habiendo dado ya el primer paso con la aprobación de la Recomendación sobre La Ética

[334] Russon, 2019.
[335] Véase RightsCon, 2019.

de la IA en su 41ª Asamblea General culminada el 24 de noviembre de 2021, ya aludida con anterioridad.[336]

Un buen punto de partida y referencia, a los efectos de crear tal Declaración, podría ser, sin duda, la Declaración Universal de Derechos Humanos aprobada en París el 10 de diciembre de 1948 por la Asamblea General de la ONU en su Resolución 217 A (III), que recoge en sus treinta artículos los derechos humanos considerados básicos, habida cuenta de los severos riesgos de vulneración de los mismos que entraña tal tecnología. No obstante, el principal obstáculo con el que toparíamos en tal caso, sería la ausencia de carácter vinculante de tal documento (siendo, sin duda, más difícil todavía si se vinculara directamente a la Declaración Universal de Derechos Humanos, que, por desgracia, no ostenta tal carácter). No obstante, la enorme amenaza que la IA representa para la humanidad sería una buena justificación y una gran oportunidad para insertar cambios y, por primera vez en la historia, aprobar en el marco de las Naciones Unidas un documento jurídicamente vinculante para todos sus miembros, con severos mecanismos de cumplimiento. Es todo cuestión de voluntad.

Otra alternativa para la adopción de una Declaración de Principios de IA a nivel internacional, con carácter vinculante, sería la del Consejo de Europa, compuesto por cuarenta y siete Estados Miembros y seis países observadores, con base en la Convención Europea de Derechos Humanos.

No obstante, siendo que tal organismo únicamente acoge aproximadamente la mitad de los países oficiales independientes del mundo, el riesgo de creación, desarrollo y uso de sistemas de IA que no cumplan con unos mínimos estándares comunes éticos y de garantía de derechos por parte de la gran multiplicidad de países que forman parte del mismo, es muy elevado y desde luego, no elimina la amenaza que supone la IA para nuestra especie.

Y a un nivel inferior, en el ámbito de la UE, la base, sin duda, sería la Carta de Derechos Fundamentales de la Unión Europea aprobada en Niza el 7 de diciembre del 2000 por el Parlamento Europeo, el Consejo de la UE y la Comisión Europea, que en virtud de lo

[336] Véase pág. 107.

dispuesto en el artículo 6 del Tratado de la UE (en adelante, TUE), tiene otorgado el mismo valor jurídico vinculante para los Estados Miembros que el que ostentan los Tratados, resultando ser fuente del Derecho de la UE. Así, el artículo 2 del TUE dispone: "*La Unión se fundamenta en los valores de respeto de la dignidad humana, libertad, democracia, igualdad, Estado de Derecho y respeto de los derechos humanos, incluidos los derechos de las personas pertenecientes a minorías. Estos valores son comunes a los Estados miembros en una sociedad caracterizada por el pluralismo, la no discriminación, la tolerancia, la justicia, la solidaridad y la igualdad entre mujeres y hombres.*", y con el fin de velar por el respeto de tales valores, el artículo 7 del TUE prevé la formas de proceder para determinar la existencia de un riesgo claro de violación grave por parte de un Estado miembro, y constatar la existencia de tal violación persistente, estableciendo la posibilidad de imponer una sanción que implica la suspensión de determinados derechos derivados de la aplicación de los Tratados al Estado miembro de que se trate, incluidos los derechos de voto del representante del Gobierno de dicho Estado miembro en el Consejo Europeo.

En relación con ello es interesante traer a colación lo dispuesto por el Consejo de la Unión Europea el 21 de octubre de 2020, fecha en que hizo públicas las conclusiones de la Presidencia alemana bajo el título "*The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change*", habiendo puesto de manifiesto que "*queremos asegurarnos de que el diseño, desarrollo, despliegue y uso de nuevas tecnologías defiendan y promuevan nuestros valores comunes y los derechos fundamentales garantizados por la Carta de los Derechos Fundamentales de la UE, al tiempo que aumenta nuestra competitividad y prosperidad*."[337]

Es importante poner de manifiesto, en tal sentido, que la Agencia de los Derechos Fundamentales de la Unión Europea, creada en 2007, con sede en Viena (Austria), juega un rol fundamental en la supervisión de la situación de los derechos fundamentales en el ámbito de la UE, recopilando, analizando, evaluando y difundiendo información y datos relativos a los mismos, y llevando a cabo investigaciones y exámenes científicos, lo cual podría ser muy útil en el ámbito de la IA.

[337] Consejo de la Unión Europea, 2020, pág. 3.

Asimismo, en caso de vulneración de cualquier derecho fundamental previsto en la Carta, existe la posibilidad de acudir al Tribunal de Justicia de la UE, que como es sabido, tiene por misión interpretar y aplicar la legislación (a través de las cuestiones prejudiciales y procedimientos de infracción), anular normas europeas (a través de recursos de anulación), garantizar que la UE actúe (a través de los recursos por omisión), y sancionar a las Instituciones europeas (a través de la acción por daños y perjuicios).

En relación con ello, las Instituciones de la UE tienen como objetivo establecer un marco legal sobre IA, que desde luego no topará con el problema de la ausencia de carácter vinculante, siendo que inevitablemente irá ligado a la Carta de Derechos Fundamentales de la Unión y podrá revestir la forma de Reglamento directamente aplicable en los Estados Miembros. No obstante, el problema en este caso, sería el de su ámbito de aplicación, que resultaría muy reducido y limitado, ya que evidentemente solo afectaría a los países de la UE.

De acuerdo con lo expuesto, a pesar de la multitud de proyectos interesantes que hay en marcha, y a pesar del enorme valor que implicaría para la humanidad, no parece que, al menos por el momento, puedan tenerse expectativas realistas sobre la creación de una Declaración Universal de Principios de IA que ostente carácter vinculante para la práctica totalidad de los países del mundo, debiéndonos conformar con meras recomendaciones o declaraciones de valores con valor orientativo que, con suerte, servirán para inspirar a los distintos gobiernos nacionales a la hora de dictar leyes que regulen la IA dentro de sus fronteras, tal y como parece que ya está sucediendo. De todas formas, tal y como se ha expuesto, sin duda estamos ante un fenómeno sin precedentes, y afortunadamente cada vez hay una mayor concienciación en los distintos ámbitos de la sociedad sobre la amenaza que supone la IA para la humanidad y la necesidad de establecer límites a la misma para garantizar que suma valor nuestra especie y no se lo resta, por lo que quizás haya sorpresas positivas.

2.4.3. PRINCIPIOS BÁSICOS

Tal y como ya se ha expuesto, el Derecho juega un rol fundamental en el ámbito de las nuevas tecnologías y, específicamente, en el de la IA. Y es que es importante reiterar que

no todo lo técnicamente posible tiene que resultar, necesariamente, ética y/o jurídicamente viable, por lo que el regulador ostenta un papel imprescindible y fundamental para establecer pautas que aseguren un uso y un desarrollo de la IA responsable, acorde con los valores y principios morales que rigen en nuestras sociedades.

Y es que no hay duda de que la IA no solo tiene un lado amable, puesto que el mismo potencial que puede servir para mejorar el mundo puede acabar virando en nuestra contra un modo muy destructivo, tal y como recoge el informe titulado "*The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*", presentado por veintiséis expertos en implicaciones de seguridad de tecnologías emergentes provenientes de un amplio rango de organizaciones y disciplinas[338].

En relación con ello, a la vista de todo lo anteriormente establecido, y teniendo en cuenta, principalmente, los derechos reconocidos en la Declaración Universal de Derechos Humanos, como base y estándar común de logro de todos los pueblos y naciones, y haciendo una recopilación/selección de aquellos valores que, de forma mayoritaria, han sido incorporados como fundamentales en las distintas estrategias, recomendaciones, declaraciones y demás iniciativas llevadas a cabo a nivel nacional e internacional, por múltiples países y organizaciones públicas y privadas, la Declaración Universal de Principios sobre IA a la que ya he hecho alusión, en mi opinión debería contener, como mínimo, los siguientes principios:

- Principio de respeto a la dignidad del ser humano, con garantía de supervisión y control, y prioridad del bienestar social y ambiental;
- Principio de respeto a la libertad y a la privacidad del ser humano, con garantía de gestión individual de los datos personales, transparencia y explicabilidad de los sistemas;
- Principio de equidad, igualdad, no discriminación del ser humano e inclusión;
- Principio de robustez, solidez técnica y seguridad; y
- Principio de responsabilidad

[338] Véase Brundage, Avin, Clark, Toner & al, 2018.

A)Principio de respeto a la dignidad del ser humano, con garantía de supervisión y control (subordinación), y prioridad de bienestar social y ambiental

Tal y como advirtió la Vicepresidenta de la Comisión Europea, Margarethe Vestager (2019-2024): "*La IA no es ni buena ni mala en sí misma, todo depende de por qué y cómo es utilizada*".[339]

Y, en relación con ello, procede poner de manifiesto que, habida cuenta del potencial riesgo que tiene la IA para la supervivencia de la especie, siendo que puede incluso llegar a dominar al ser humano, tal y como ya se ha anunciado en puntos anteriores, es necesario asegurarse que tal tecnología se diseña, se desarrolla y se utiliza en beneficio del ser humano, que debe ostentar el papel central en el ámbito de la misma, con garantía de pleno respeto fundamentalmente a su derecho a la dignidad.

En referencia a ello, es interesante hacer alusión a algunos de los programas y proyectos más relevantes desarrollados en el ámbito de la creación de una IA centrada en el ser humano como son, por un lado, el "*MIT-Human Centered AI*", un conjunto de investigaciones y cursos del MIT (Massachussetts, EEUU) centrados en el diseño, el desarrollo y el despliegue de sistemas de IA que aprendan y colaboren con los humanos de una manera profunda y significativa con dos objetivos: la mejora contínua de los sistemas de IA mediante el aprendizaje a través de los seres humanos; y la creación de una experiencia de interacción humanos-robots efectiva y satisfactoria[340]; y, por otro lado, el Stanford Institute for Human-Centered AI de la Universidad de Stanford (California, EEUU), cuyos codirectores Fei-Fei Li y John Etchemendy afirman: "*Para que la IA sirva a las necesidades colectivas de la humanidad debe incorporar una comprensión de lo que nos mueven: física, intelectual y emocionalmente. Es fundamental que diseñemos IA que pueda comprender el lenguaje humano, los sentimientos, las intenciones y los comportamientos, e interactuar con matices y en múltiples dimensiones. Para lograr esto, los creadores y diseñadores de la IA deben ser ampliamente representativos de la*

[339] Masdeu, 2020.
[340] Massachussetts Institute of Technology, 2017.

humanidad. Esto requiere una verdadera diversidad de pensamiento: género, etnia, nacionalidad, cultura y edad, así como en todas las disciplinas."[341]

Y es que para conseguir una auténtica IA centrada en el interés del ser humano, es imprescindible transmitir a los sistemas valores humanos y conseguir que estos los entiendan y actúen de acuerdo con ellos. Pero no solo eso, sino que resulta, asimismo, fundamental que los humanos comprendamos cómo funcionan los sistemas de IA y sepamos qué hay detrás de ellos, para poder así lograr una relación máquina-humano fluida y de confianza que permita crear y construir en una misma dirección. Y ello, desde luego, es posible, no solo porque, tal y como se ha avanzado, los sistemas de IA son neutros por naturaleza y, por ende, se diseñan y se desarrollan en función de las decisiones/directrices fijadas por el que lleva a cabo tales tareas (que en virtud de lo expuesto, debería ser un ser humano), sino porque además, si existe un proyecto responsable y basado en el interés humano, es técnicamente posible que los sistemas puedan ser programados para que resulte viable rastrear, comprender y controlar sus procesos de toma de decisiones por parte de los seres humanos.

Y es que, si pretendemos que la IA sea empleada en beneficio de nuestra especie y no en perjuicio de la misma, resulta fundamental sentar las bases para que esta se convierta en el centro de tal tecnología, y para que el derecho a la dignidad de la persona sea respetado en todo caso, ya que es el derecho que, en mi opinión, más nos define e identifica como especie.

Según la Agencia de la ONU para los refugiados (UHNCR-ACNUR) "*La dignidad humana es el derecho que tenemos todos los seres humanos a ser valorados como sujetos individuales y sociales, con nuestras características particulares, por el simple hecho de ser personas. La dignidad supone, además, el derecho a ser nosotros mismos y a sentirnos realizados, lo que se manifiesta en la posibilidad de elegir una profesión, expresar nuestras ideas y respetar a los demás*." Y añade "*Se oponen a la dignidad aspectos como los tratos humillantes, la discriminación en todas sus facetas o la desigualdad*."[342]

[341] Stanford University, 2019.
[342] ACNUR, 2018.

Tal y como manifiesta Miguel L. Lacruz Mantecón, profesor de la Universidad de Zaragoza, lo que nos define como personas, seres racionales, desde luego, no es la inteligencia (entendida como la capacidad para resolver problemas), puesto que, por un lado, la mayoría de animales también cuentan con tal aptitud y, por otro lado, las personas con inteligencia disminuida o anulada por una enfermedad no dejan de ser consideradas como tales. Y es que según concluye "*es nuestra pertenencia a la especie humana, únicamente este hecho, el que nos confiere la dignidad de humanos y por tanto conciencia (otros dirán alma), y por ello el sujeto carente de toda inteligencia, el enfermo en coma, sigue siendo un ser humano*."[343]

Así, los sistemas de IA deben en todo caso respetar el derecho fundamental a la dignidad humana y ello, bajo mi punto de vista, únicamente puede garantizarse si conseguimos que nuestra especie no pierda el control de tal tecnología, puesto que, desde luego, una vez salga de nuestro "ámbito de poder" su supervisión devendrá prácticamente inabarcable. Y es que, ya se ha puesto de manifiesto en puntos anteriores que los sistemas de IA son cada vez más autónomos y sofisticados y que, muchos de ellos, toman ya decisiones bajo criterios que resultan prácticamente imposibles de descifrar por los técnicos, lo que ha hecho correr ríos de tinta en relación con la denominada caja negra o *black box*, a la que se hará especial mención más adelante.

Y en tal sentido, resulta interesante un artículo escrito por Daron Acemoglu y Pascual Restrepo[344] que plantea que quizás se está avanzando en la dirección incorrecta, poniendo demasiado empeño en conseguir una mayor autonomía y automatización de los sistemas cuando en realidad interesa más trabajar para que la IA se convierta en un complemento (no sustituto) beneficioso para nuestras sociedades.

Para llevar a cabo ese control de los sistemas de IA, no obstante, resulta imprescindible establecer la obligación de que estos siempre puedan sean supervisados, en último término, por un ser humano, siendo que además ese es el único escenario capaz de generar un clima de verdadera confianza entre los ciudadanos hacia tal tecnología, lo que resulta

[343] Rogel, Lacruz, Mozo & Díaz, 2018.
[344] Acemoglu & Restrepo, 2018.

fundamental para su implementación en beneficio de la sociedad, habida cuenta del recelo con el que esta es mirada en la actualidad.

Y todo ello, siendo que implica el objetivo de garantizar y fomentar el interés del ser humano, debe redundar sin duda en el bienestar colectivo, tanto social como ambiental.

Yes que la IA, tal y como ya se ha venido apuntando reiteradamente a lo largo de estas páginas, tiene un potencial enorme para influir de forma muy positiva en innumerables ámbitos de la sociedad (trabajo, transporte-logística, sanidad, industria, servicios, justicia, defensa, etc) y, por ende, un buen enfoque y un correcto uso de la misma permitiría aprovechar multiplicidad de oportunidades que nos ayudarían a hacer más fácil y eficiente nuestro día a día como individuos y como sociedad.

Uno de esos ámbitos en los que la IA puede generar un mayor impacto, por los efectos tan globales que implica es, sin duda, el medioambiental. Y es que tal impacto, en caso de no ser regulado y limitado, puede ser muy perjudicial, habida cuenta de que los sistemas de IA suelen requerir un gran consumo de energía, por lo que resulta necesario avanzar en consonancia con la normativa y las políticas energéticas y de medio ambiente que aplican a nivel nacional e internacional para dar buenos pasos en esa dirección. No obstante, cada vez más investigadores aseguran que el *big data* y la IA pueden ser altamente útiles justamente para mejorar la eficiencia energética, renovar industrias (entre otras, la agricultura), encontrar nuevos materiales de construcción sostenibles o *eco-friendly*, etc, lo que sin duda, resulta un camino interesantísimo a explorar, ya que con un buen equilibrio entre el riesgo y el beneficio se podrían conseguir resultados muy alentadores.

En definitiva, una IA centrada en el interés humano (individual y colectivo), únicamente puede conseguirse mediante la creación, el desarrollo y el uso de sistemas respetuosos con el derecho a la dignidad del mismo, principalmente, siendo que del mismo derivan el resto de derechos humanos que tenemos reconocidos como especie, y con garantía de posibilidad de supervisión y control humano de tal tecnología, de forma última, siempre con el foco puesto en el bienestar social y ambiental.

B) Principio de respeto a la libertad y la privacidad del ser humano, con garantía de gestión individual de datos personales, transparencia y explicabilidad de los sistemas

Son constantes (y crecientes) los impactos que la IA tiene en nuestras vidas y en las decisiones que tomamos desde hace ya algún tiempo. Así, no resulta ajeno a nadie recibir ofertas por correo electrónico de productos o servicios relacionados con sus gustos, hallar sugerencias de películas o series que les pueden encajar o ver aparecer imágenes en la pantalla del ordenador sobre viajes que han estado planeando pero todavía no han decidido contratar, etc. Y ello, si bien en ocasiones puede resultar útil, esconde una posible y voraz amenaza a nuestro derecho a la libertad y a la privacidad.

Entre otras, la Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de IA y robótica, en su Considerando L otorga una especial relevancia a los derechos de protección de datos, privacidad y seguridad, al establecer: "*Considerando que el marco europeo para la Inteligencia Artificial ha de desarrollarse sobre la base del pleno respeto de los derechos consagrados en la Carta de los Derechos Fundamentales de la Unión Europea, y en particular de los principios de protección de datos, privacidad y seguridad*."

Y es que ¿de qué forma un algoritmo ha obtenido y tratado información relativa a mis gustos o preferencias y ha determinado qué ofertas o promociones mandarme? ¿he dado yo consentimiento para que se traten mis datos? ¿tengo verdadera libertad de elegir cuando no paro de recibir mensajes subliminales que no ha solicitado?.

Ello, no obstante, es solo la punta del *iceberg*, y desde luego los impactos que la IA puede tener en la libertad y la privacidad de los ciudadanos van mucho más allá de las decisiones relativas al consumo.

Generalmente se dice que "el diablo está en los datos". Y es que, los sistemas de IA básicamente se nutren de datos, lo cual puede, en ocasiones, resultar sumamente peligroso, ya que estos pueden haberse obtenido de forma ilícita y/o sin consentimiento de sus titulares, con la consiguiente posible vulneración del derecho a la intimidad, a la dignidad, a la libertad y, en ocasiones, a la propiedad; pueden haberse captado de forma sesgada o selectiva, lo cual podría implicar una infracción del derecho a la igualdad y a la no discriminación; pueden resultar opacos, con la posibilidad de vulnerar el derecho a la

información, entre otros, por lo que hay que prestar muchísima atención y, sin duda, adoptar aquellas medidas que garanticen la legalidad de los mismos.

En relación con ello, los investigadores, por un lado, están trabajando para que sean los propios algoritmos los que detecten y corrijan tales problemáticas en caso de que existan, si bien aun queda mucho camino por recorrer; y por otro lado, se está luchando por parte de todos los sectores implicados para conseguir que los sistemas de IA se programen, ya de inicio, con datos "de calidad", es decir, legalmente obtenidos, representativos de todos los sectores de la población, sin sesgos, transparentes, etc. Asimismo, y dado que los sistemas de *Deep Learning* (los más extendidos en la actualidad) requieren grandes cantidades de datos, una de las mayores prioridades de los investigadores es conseguir reducir la dependencia de estos

Y es que el uso de datos "de calidad" en los sistemas de IA es, sin duda, la forma (al menos en la actualidad) más eficiente para garantizar buenos resultados y para asegurar que esta actúa y se emplea en beneficio del ser humano y de la sociedad, con respeto a sus derechos, puesto que, como se ha dicho, múltiples de los derechos humanos reconocidos pueden resultar vulnerados como consecuencia del uso de datos de "mala calidad" para nutrir y/o entrenar sistemas de *Machine Learning*.

En tal sentido, para poder calificar la información como "datos de calidad", la Agencia Europea de Derechos Fundamentales ha establecido dos cuestiones a evitar:

- ✓ errores de representación, lo que implica que la información no cubre la totalidad de la población que debería cubrir; y
- ✓ errores de medición, lo que implica que los datos no miden lo que originariamente estaba planeado. [345]

El Comisario de Derechos Humanos del Consejo de Europa, en relación con ello, en el informe titulado "*Unboxing Artificial Intelligence:10 steps to protect Human Rights*", puso de manifiesto que "*El procesamiento de datos en el contexto de los sistemas de IA debe ser*

[345] Agencia Europea de Derechos Fundamentales, 2019, págs. 11-12.

proporcionado en relación con el propósito legítimo perseguido a través de dicho procesamiento, y en todas las etapas del procesamiento debe reflejar un equilibrio justo entre los intereses perseguidos a través del desarrollo y el despliegue del sistema de IA y los derechos y libertades en juego"[346] y, además, sugirió a los Estados Miembros la implementación de modo efectivo de la "*Convention for the protection of individuals with regard to the processing of personal data*" (Convention 108+)[347], así como cualquier otro instrumento nacional o internacional sobre protección de datos y privacidad que resulte aplicable.

Como ejemplo de malas prácticas en el ámbito de la captación y la gestión de los datos, es interesante hacer mención al denominado caso Cambridge Analytica, una empresa de consultoría británica que ofrecía entre sus servicios el análisis de datos y el cambio de comportamientos sociales a través de la IA, que fue contratada en 2016 para las campañas de dos candidatos del Partido Republicano de EEUU, primero Ted Cruz y luego Donald Trump, con la promesa de identificar las personalidades de los votantes estadounidenses e influir en su comportamiento para lo que, al parecer, empleó datos de Facebook captados masivamente y sin consentimiento de sus titulares (más de cincuenta millones), tal y como desvelaron los periódicos The New York Times (EEUU) y The Observer (Reino Unido). Como consecuencia de ello, la empresa se liquidó y la Comisión Federal de Comercio de Estados Unidos condenó a la antedicha red social a pagar la mayor sanción jamás impuesta a una compañía por vulnerar la privacidad de sus clientes, una multa de cinco mil millones de dólares, por las malas prácticas en la gestión de la seguridad de los datos de ochenta y siete millones de usuarios y a crear un comité independiente para asuntos de privacidad, que deberá quedar fuera del control de su fundador y Consejero Delegado Mark Zuckerberg.[348]

Respecto de los datos personales, es un hecho evidente y no controvertido que EEUU y China fueron los primeros países en llevar a cabo ingentes inversiones y proyectos de captación masiva de datos de tal clase, lo que les han generado enormes beneficios, oportunidades y rentabilidades, tanto en el sector público como en el sector privado y, por

[346] Consejo de Europa, 2019, págs. 11-12.
[347] Véase Consejo de Europa, 2018.
[348] Véase Rosenberg, Confessore & Cadwalldar, 2018.

ende, puede asegurarse que, en tal sentido, "han ganado la batalla". De hecho, Kai-Fu Lee, autor del libro "*AI Superpowers*", ha asegurado "*si los datos son el nuevo petróleo, China es la nueva OPEP*"[349]. La UE, en cambio, ha quedado "a la cola" de tal carrera, quizás por su falta de estrategia política en tal sentido o quizás por los elevados estándares de derechos y libertades por los que se rige, que desde luego juegan un rol fundamental como garantía de protección para los ciudadanos, pero operan como límite en tal ámbito.

No obstante, la UE ha reaccionado y pretende subsanar ese decalaje imponiéndose a la cabeza de la nueva era, la de los datos industriales (generados por las empresas y los organismos públicos), y es que la Comisión Europea considera que la Unión posee una enorme base de datos de tal clase, lo que supone una ventaja frente al resto de países que hay que aprovechar, siendo su principal objetivo el de crear un "mercado único europeo" donde los datos personales y no personales puedan ser utilizados por las empresas y el sector público para crear valor e innovar, siempre con pleno respeto de los valores y derechos en los que se inspira la UE.[350]

En la actualidad, la UE planea lanzar una "estrategia europea de datos" que busca convertir la Unión en líder de una sociedad impulsada por estos, con la creación de un mercado único que permitirá que fluyan libremente por la UE, en beneficio de las empresas, los investigadores y las Administraciones Públicas. Y es que se considera que las personas, las empresas y las organizaciones deben estar capacitadas para tomar mejores decisiones a partir del conocimiento que aportan los datos no personales, que deben estar a disposición de todos.[351]

Y es que en el ámbito de la UE, tal y como afirma la Presidenta de la Comisión Europea, Ursula von der Leyen (2019-2024), resulta imprescindible crear un clima de confianza de los ciudadanos hacia la tecnología, y en especial hacia la IA: "*Queremos que los ciudadanos confíen en la tecnología nueva. La tecnología es siempre neutral. Todo depende de lo que hagamos con ella. Por lo tanto queremos que la aplicación de estas nuevas tecnologías merezcan la confianza de nuestros ciudadanos. Por eso promovemos*

[349] Asia House, 2018.
[350] Véase Masdeu, 2020.
[351] Véase Unión Europea, 2019.

un enfoque para la IA centrado en los humanos y responsable. Tenemos que asegurar que los datos que se utilizan están libres de sesgo".[352] No obstante, para que la UE pueda realmente cumplir con sus objetivos y se convierta en abanderada de la implementación y el desarrollo de la IA, debe seguir aumentado sus inversiones ya que, sin duda, uno de los puntos clave para el avance y el desarrollo es la investigación, y sin embargo, existe un gran riesgo de "fuga de cerebros" por falta de recursos, tal y como se subraya en la Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de IA y robótica: "*las remuneraciones de los investigadores de la Unión siguen siendo muy inferiores a las de sus homólogos de los Estados Unidos y China, reconociéndose que esa es la principal razón que les lleva a abandonar Europa*."[353]

En relación con ello, suele ponerse de relieve la antedicha ventaja competitiva que ostenta especialmente China en el ámbito de la IA, como consecuencia principalmente de la gran cantidad de datos personales que maneja, habida cuenta de su enorme población y la prácticamente ausente limitación normativa en el uso de sistemas de vigilancia gubernamental masiva. No obstante, en mi opinión, la verdadera ventaja estratégica y competitiva en el desarrollo de sistemas de IA en el futuro no va a radicar en la cantidad de datos que se manejen, sino en la calidad de los mismos, que es donde reside la clave. Y es que está claro que un sistema de aprendizaje automático entrenado con rostros occidentales, por ejemplo, no funciona bien con rostros orientales. Y así lo demuestra, entre otros, el experimento llevado a cabo por Vicomtech (un centro tecnológico especializado en IA, ubicado en el País Vasco), consistente en el desarrollo de un sistema de IA para la detección del sueño de los conductores de vehículos a través del reconocimiento facial, en que se emplearon grandes cantidades de datos biométricos de personas occidentales, habiendo conseguido un gran porcentaje de éxito. No obstante, al ser presentado tal proyecto y ser probado por una persona de raza oriental, falló al detectar que esta estaba dormida, cuando en realidad permanecía despierta, como consecuencia de la forma achinada de sus ojos, tal y como explica Doña Oihana Otaegui Madurga, Directora de Sistemas de Transporte Inteligentes e Ingeniería de la antedicha empresa, lo que, sin duda, es un claro ejemplo de cómo la calidad de los datos marca el éxito de los sistemas de IA que los emplean.

[352] Valdeolmillos, 2020.
[353] Punto 21.

Y así se recoge en la Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de IA y robótica, que: "38. *Recuerda que disponer de datos de alta calidad que además sean significativos es esencial para una auténtica competitividad en el sector de la inteligencia artificial, y pide a los poderes públicos que garanticen formas de producir, compartir y regular los datos públicos convirtiéndolo en un bien común, al tiempo que se preservan la privacidad y los datos sensibles"*; y "*39. Subraya la importancia de la calidad de los datos utilizados en el aprendizaje profundo; señala que el uso de datos de baja calidad, anticuados, incompletos o incorrectos puede dar lugar a malas previsiones y a su vez a discriminaciones y sesgos*."

En relación con ello, procede hacer referencia a las regulaciones que, en el ámbito de la protección de datos, y con el ánimo de garantizar el derecho a la libertad y a la privacidad de los ciudadanos, principalmente, han adoptado tanto la Unión Europea como España, por resultar las que nos resultan aplicables.

En tal sentido, debe ponerse de relieve que el artículo 8 de la Carta de Derechos Fundamentales de la UE reconoce el derecho de los ciudadanos de la Unión a que protejan sus datos personales, el punto de referencia es sin duda el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de tales datos, denominado Reglamento general de protección de datos (RGPD), por el que se derogó la Directiva 95/46/CE.

No obstante, a largo de los últimos años se han ido aprobando y publicando diversas resoluciones y/o proyectos europeos relativos al ámbito de los datos. Entre otros, cabe destacar la Directiva (UE) 2016/680, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales relacionados con infracciones penales o con la ejecución de sanciones penales, y a la libre circulación de tales datos, que entró en vigor el 5 de mayo de tal año y establece una específica protección del derecho fundamental a la protección de datos cuando estos sean empleados por las autoridades policiales y judiciales.

Asimismo, es interesante hacer alusión a la Comunicación de la Comisión de 19 de abril de 2016 (pocos días antes de la publicación del RGPD) titulada "*Iniciativa Europea de Computación en la Nube: construir en Europa una economía competitiva de los datos y del conocimiento*"[354], que hace un llamamiento al establecimiento de una infraestructura de datos europea basada en capacidades de informática de alto rendimiento de vanguardia y al desarrollo de un completo ecosistema de informática de alto rendimiento capaz de desarrollar nuevas tecnologías europeas y hacer realidad los superordenadores; así como la Comunicación de la Comisión de 10 de mayo de 2017, relativa a la revisión intermedia de la aplicación de la "*Estrategia para el Mercado Único Digital. Un mercado único digital conectado para todos*" [355] que ve en la informática de alto rendimiento un elemento decisivo para la digitalización de la industria y la economía de los datos.

Asimismo, el 14 de noviembre del 2018 se aprobó el Reglamento (UE) 2018/1807 del Parlamento Europeo y del Consejo, relativo a un marco para la libre circulación de datos no personales en la Unión Europea, y el 20 de junio del 2019 se dictó la Directiva (UE) 2019/1024 del Parlamento Europeo y del Consejo, relativa a los datos abiertos y la reutilización de la información del sector público.

Desde el 2018, además, existe el Comité Europeo de Protección de Datos (CEPD)[356], formado por un presidente y dos vicepresidentes (designados para un mandato de cinco años renovable), todas las autoridades nacionales de protección de datos y el Supervisor Europeo de Protección de Datos (SEPD)[357], que se encarga de garantizar que, a la hora de tratar datos personales, las instituciones y los organismos de la UE respeten el derecho a la intimidad de los ciudadanos. El CEPD tiene como objetivo asegurar que el RGPD y la Directiva sobre protección de datos en el ámbito penal se apliquen de manera coherente en los países de la UE, así como en Noruega, Lichtenstein e Islandia.

En el seno de la Comisión Europea existe el "Responsable de la Protección de Datos" (RPD) que garantiza, de forma independiente, que tal institución aplique correctamente la

[354] Véase Comisión Europea, 2016.
[355] Véase Comisión Europea, 2017.
[356] Véase Unión Europea, 2018.
[357] Véase Unión Europea, 2004.

normativa en materia de protección de los datos personales de los particulares, con la llevanza incluso de un registro en que se publican todas las operaciones de la Comisión que implican el tratamiento de datos de tal clase.[358]

Para el año 2025 se prevé un 530 % de incremento del volumen global europeo de datos (de 33 zetabytes en 2018 a 175 zetabytes), ochocientos treinta mil millones de euros de valor de la economía de los datos en la EU27, frente a los trescientos mil millones de euros (2,4 % del PIB de la UE) que tenía en 2018; once millones de profesionales de los datos en la EU27, frente a los seis millones que había en 2018; y un 65 % de población de la UE con competencias digitales básicas, frente al 57 % que existía en 2018.[359]

En el Consejo de Europa, por su parte, ya el 23 de enero del 2017 el Comité Consultivo del Convenio 108 publicó el informe denominado "*Guidelines on the Protection Of Individuals With Regard To The Processing Of Personal Data In A World Of Big Data*"[360], habida cuenta de la necesidad de garantizar la protección de la autonomía personal basada en el derecho de una persona a controlar sus datos personales y el procesamiento de los mismos. Posteriormente, el 25 de enero del 2019, asimismo, publicó el documento titulado "*Guidelines on AI and Data Protection*" [361] que puso de manifiesto las posibles consecuencias adversas para las personas y la sociedad que puede tener la IA y, con el fin de evitarlo, estableció que las Partes del Convenio 108 tenían por objeto garantizar y permitir un desarrollo y un uso de la IA que respeten los derechos a la privacidad y a la protección de datos (en aplicación del artículo 8 del Convenio Europeo de Derechos Humanos), mejorando así los derechos humanos y las libertades fundamentales. Para ello se proporcionó un conjunto de pautas y medidas de referencia que los gobiernos, los diseñadores de IA, los fabricantes y los proveedores de servicios deberían seguir para garantizar que las aplicaciones de tal tecnología no menoscaben la dignidad y el resto de derechos humanos y libertades fundamentales de cada individuo y, en particular, el derecho a la protección de datos.

[358] Véase Comisión Europea, 2018.
[359] Comisión Europea, 2019.
[360] Véase Consejo de Europa, 2017.
[361] Véase Consejo de Europa, 2019.

En España, el punto de referencia de la regulación en materia de protección de datos es la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD), que vino a sustituir a la Ley Orgánica 15/1999, de Protección de Datos de Carácter Personal y cuyo objetivo, principalmente, es la protección de la integridad, la intimidad y la privacidad del individuo, en virtud de lo dispuesto en el artículo 18 de la Constitución Española.

En relación con ello, existe en nuestro país la Agencia Española de Protección de Datos, la autoridad pública independiente, en virtud de lo previsto en la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, encargada de velar por la privacidad y la protección de los datos de los ciudadanos de España. Asimismo, existen dos Agencias autonómicas de Protección de Datos: la Autoridad Catalana de Protección de Datos[362] y la Agencia Vasca de Protección de Datos[363].

No obstante, en el ámbito de la investigación criminal, que es el que ocupa la presente tesis doctoral, resulta de aplicación una ley especial, la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, que entró en vigor el 16 de junio de 2021 (excepto su Capítulo IV, que entra en vigor el 16 de diciembre del mismo año).

Una vez expuesto lo anterior, procede poner de manifiesto que para analizar qué derechos pueden verse amenazados por la aplicación de la IA es evidente que resulta fundamental poder conocer, con absoluta transparencia y de una forma comprensible, la base sobre la que los sistemas que emplean tal tecnología toman las decisiones que nos afectan. Y es que, para poder detectar la vulneración de un derecho es imprescindible conocer previamente el contenido del proceso decisorio y su funcionamiento, lo que además deviene fundamental para que los sistemas de IA se ganen la confianza de los ciudadanos, absolutamente necesaria para que su aplicación resulte realmente útil y eficiente.

[362] Véase APDCAT, s.f..
[363] Véase AVPD, s.f..

Bajo mi punto de vista, uno de los focos de riesgo más importantes a los que tiene que hacer frente la ciencia jurídica para asegurar un impacto positivo de la IA en la sociedad es, sin duda, la opacidad de los sistemas, que hace extremadamente difícil (y en ciertos casos, imposible) evaluar la viabilidad jurídica de los mismos. Y es que resulta evidente que si no se garantiza a los usuarios la transparencia y explicabilidad del contenido de los mecanismos que se les aplican, estos no van a ser capaces de otorgar su confianza a dicha tecnología.

No obstante, y a pesar de que durante los últimos años se ha extendido el uso de algoritmos opacos, sobre todo en el ámbito del sector privado, lo cierto es que en la actualidad está proliferando una conveniente conciencia sobre la necesidad de que estos sean transparentes y explicables.

El principal problema que nos encontramos para controlar qué tipo de datos están siendo empleados por los sistemas de IA y de qué forma están siendo tratados es que en estos suele siempre existir una caja negra o *black box* que dificulta o impide acceder a su contenido, siendo que los algoritmos interiorizan y emplean datos de formas difíciles de controlar y entender por los humanos.

El problema de la caja negra o *black box* puede ser definido como "*la incapacidad de comprender completamente el proceso de toma de decisiones de un sistema de IA y la incapacidad de predecir las decisiones o los resultados del mismo*."[364]

Y en relación con ello, por ejemplo, el campeón mundial de ajedrez Garry Kasparov sugirió en su día, al ser vencido por el programa de IBM Deep Blue, que este podía haber hecho trampas, habida cuenta de que no entendía alguno de sus movimientos y el contenido y modo de funcionar del algoritmo permanecía oculto, existiendo una caja negra o *black box* que dio lugar a sus serias sospechas.[365]

Dicha problemática en los sistemas de IA suele clasificarse en:

[364] Bathaee, 2018, pág. 891.
[365] Véase Chandrasekaran, 1997.

a) *cajas negras <u>fuertes</u>* (*strong black boxes*),

que hacen referencia a aquellos procesos de toma de decisiones que emplean IA y que son completamente opacos para los humanos, no existiendo forma de determinar ni cuál ha sido el proceso que ha llevado al sistema a tomar una determinada decisión o a hacer cierta predicción, ni qué información ha resultado determinante en la toma de la decisión, ni tampoco permiten obtener una clasificación, por orden de importancia, de las variables procesadas por el sistema de IA.

En tales casos, debe ponerse de manifiesto que ni siquiera puede realizarse un análisis *a posteriori* mediante ingeniería inversa para obtener tales informaciones;

b) *cajas negras <u>débiles</u>* (*weak black boxes*),

que hacen referencia a aquellos procesos de toma de decisiones que emplean IA y que, si bien también resultan opacos para los humanos, pueden ser sometidos a un análisis *a posteriori* de ingeniería inversa, o incluso analizarse para conseguir una clasificación flexible, por orden de importancia, de las variables que el sistema de IA tiene o ha tenido en cuenta para tomar una determinada decisión o hacer cierta previsión.[366]

La falta de transparencia puede surgir, por un lado, por la "complejidad", habida cuenta de la dificultad o la complicación de la estructura de un algoritmo, como en el caso de una red neuronal profunda, en que existen miles de neuronas artificiales que trabajan juntas de modo difuso para resolver un problema; y por otro lado, por la "dimensionalidad" en aquellos casos en que la IA utiliza un algoritmo de aprendizaje automático (*Machine Learning*) que se basa en relaciones geométricas que los humanos no pueden visualizar.[367]

La explicabilidad hace referencia a la posibilidad de dar a conocer y hacer entender, de forma clara y comprensible, el modo de funcionar de los sistemas de IA. En tal sentido, existen dos perspectivas de dicha característica: por un lado, la abstracta, que consiste en la explicación de la forma en que un sistema toma decisiones, mediante la declaración y la

[366] Bathaee, 2018, págs. 905-906.
[367] Bathaee, 2018, pág. 901.

publicación de las reglas que este sigue, sin hacer referencia a ninguna decisión específica; y, por otro lado, la concreta, que consiste en la explicación del proceso de adopción de una determinada decisión genérica, y requiere la declaración o publicación de las razones o justificaciones empleadas para un resultado en particular, en lugar de una descripción del proceso de toma de decisiones.[368]

Uno de los principales problemas con que nos topamos en el ámbito de la explicabilidad, tal y como he podido deducir de diversas conversaciones con ingenieros y técnicos de IA, es la diferencia de puntos de vista entre los diseñadores de los sistemas de IA y los ciudadanos. Así, los primeros preguntan ¿qué información exacta se necesita obtener para dar por cumplido el principio de transparencia? Y ante tal pregunta, la respuesta (al menos mía, lega en conocimientos tecnológicos), siempre ha sido "No lo sé, decidme vosotros qué información hay y cuál me podéis proporcionar". Y es que el más puro desconocimiento de cómo funciona un sistema de IA hace que al recibir tal pregunta vengan a la cabeza multitud de ideas desordenadas y abrumadoras que no permiten lanzar una respuesta clara. No obstante, tras reflexionar y leer sobre el asunto, entiendo que el objetivo mínimo y principal que debe perseguir toda explicación de un sistema de IA es el de permitir a un observador externo entender y determinar en qué medida un factor particular ha sido determinante o influyente en la toma de decisiones y qué información ha sido empleada para ello.

Tal cuestión no es una tarea nada fácil, máxime si atendemos a la enorme complejidad que alberga el funcionamiento de muchos sistemas de IA, lo que dificulta mucho la traducción y posterior explicación en un lenguaje fácil y comprensible para todos aquellos legos en la materia. Para ello, de las conversaciones con expertos de diversos ámbitos del sector, he sacado la conclusión de que resultaría útil la existencia de una figura intermedia, entre los diseñadores e investigadores de los sistemas de IA y el resto de la sociedad, que ayudara a poner en común inquietudes y posibilidades técnicas, y jugara un papel de "traductor tecnológico", que en el ámbito del proceso penal podría actuar como perito.

[368] Doshi-Velez & Kortz, 2017, pág. 2.

En relación con todo ello, Danielle Keats Citron, Profesora de Derecho de la Universidad de Boston (EEUU), fue una de las primeras juristas en reclamar públicamente, entre otras, la realización de auditorías y pruebas de los sistemas de IA, la publicación de explicaciones detalladas sobre el contenido y el funcionamiento de estos, la introducción de planes de educación para los usuarios sobre la falibilidad de máquinas, y la proliferación de códigos de acceso público.[369]

Sin embargo, en la actualidad, millones de sistemas de IA siguen teniendo un fondo opaco y absolutamente desconocido no solo para los usuarios sino también, tal y como ya he anunciado, para sus propios creadores. En relación con ello, no obstante, resulta muy preocupante saber que en ocasiones se diseñan sistemas de IA que contienen *black box* cuando en realidad no habría necesidad, ya que podrían hacer el mismo papel los sistemas interpretables. Y esa es la conclusión a la que llegaron Cynthia Rudin, profesora de informática de la Universidad de Duke (Carolina del Norte, EEUU) y Joanna Radin, Profesora de historia de la medicina en la Universidad de Yale (Connecticut, EEUU), tras la celebración en Monreal (Canadá) en el año 2018 de la conferencia anual "*Neural Information Processing Systems (NeurIPS) conference*" (organizada por Google, Fair Isaac Corporation (FICO) y académicos de las universidades de Berkeley, Oxford, Imperial, UC Irvine y MIT), donde se llevó a cabo el reto-competición denominado "*Explainable Machine Learning Challenge*", cuyo objetivo era la creación, por parte de diversos equipos de expertos, de un modelo de caja negra o *black box* complicado y la posterior explicación a la audiencia de cómo funcionaba. No obstante, uno de los equipos inscritos en la competición no siguió las reglas y, en lugar de presentar un modelo con caja negra, creó un modelo completamente explicable que arrojaba los mismos resultados que el opaco, lo que puso en evidencia que el *black box* se emplea incluso cuando no es necesario.

En relación con ello, las antedichas profesionales pusieron de manifiesto que "*la creencia de que la precisión debe sacrificarse por la interpretabilidad es inexacta. No obstante, esta ha permitido a las empresas comercializar y vender modelos patentados o complicados con cajas negras para la toma de decisiones de alto riesgo cuando, en realidad, existen modelos interpretables muy simples capaces de realizar las mismas tareas. Así, se permite*

[369] Véase Citron, 2007, págs. 1.013-1.305.

a los creadores de modelos de IA obtener ganancias sin tener en cuenta las consecuencias perjudiciales para las personas afectadas. Pocos cuestionan estos modelos porque sus diseñadores afirman que los modelos deben ser complicados para ser precisos. (…) *Insistamos en que no se utilicen modelos de aprendizaje automático con caja negra para decisiones de alto riesgo a menos que no se pueda construir un modelo interpretable que logre el mismo nivel de precisión. Es posible que siempre se pueda construir un modelo interpretable, simplemente el problema es que no lo hemos intentado. Quizás si lo hiciéramos, nunca usaríamos cajas negras para estas decisiones de alto riesgo*."[370]

Y, asimismo, también resulta ciertamente alarmante otra de las cosas que ocurrieron en la antedicha conferencia. Así, antes de anunciar a los ganadores del desafío, se pidió a la audiencia, compuesta principalmente por expertos en los ámbitos de las finanzas, la robótica y el aprendizaje automático o *Machine Learning*, que participaran en un experimento. En tal contexto, se solicitó a los presentes que se imaginaran que tenían cáncer y que necesitaban cirugía para extirpar un tumor maligno. Tras ello, se mostraron dos imágenes en la pantalla: en una se mostraba a un cirujano humano, al que se podían hacer preguntas, pedir explicaciones sobre la cirugía, hacer advertencias etc, y tenía un 15% de posibilidades de fallo mortal durante la cirugía; en la otra, se mostraba un brazo robótico con el que no se podía tener ningún tipo de interacción previa a la cirugía, pero tenía solo un 2% de probabilidad de fallo mortal durante la misma. Así, el robot estaba destinado a simular una caja negra de IA y requería confianza total, puesto que tal y como se ha dicho, no admitía preguntas, advertencias, ni ofrecía explicaciones.

Tras ello, no obstante, al solicitar a la audiencia que votara cuál de las dos opciones de cirugía escogerían, todos menos uno votó por el robot, lo cual resulta altamente peligroso e incomprensible, sin duda, y demuestra que los ciudadanos asociamos IA a precisión, muchas veces a cualquier precio, y creemos que los resultados que nos anuncian van a beneficiarnos sin plantearnos, ni siquiera, qué hay detrás del sistema (qué datos se han empleado, para qué parte de la población se ha establecido la estadística de éxito, etc), lo cual es muy peligroso.

[370] Rudin & Radin, 2019, pág. 2.

Los principios de transparencia y explicabilidad de los sistemas de IA son fundamentales en cualquier ámbito de aplicación, si bien cobran, sin duda, especial relevancia en el ámbito del sector público, puesto que debe garantizarse que estos se emplean en interés colectivo y con máximo respeto a los derechos de los ciudadanos, habida cuenta, principalmente, de que las consecuencias del uso de estos suelen ser inevitables, ya que los ciudadanos no eligen si quieren que les sean aplicados o no, quedan sometidos a ellos "y punto". Y es que, en el ámbito del sector privado, si el cliente sabe que se está aplicando por una determinada empresa un algoritmo que resulta opaco, puede optar por contratar con la competencia, pero en el caso del sector público no se concede la opción de elegir sobre su aplicación, por lo que deben asegurarse las máximas garantías.

A modo de ejemplo, es interesante hacer especial referencia al conocido "caso SCHUFA", en Alemania, y al denominado "caso Parcoursup", que generaron un profundo clima de malestar y desconfianza en Alemania y Francia, respectivamente.

Por un lado, ya en 2012, en Alemania se causó un gran revuelo por el mencionado caso SCHUFA, que surgió tras unas publicaciones del periódico Die Spiegel[371] que reveló el plan de dicha agencia de crédito (la más grande del país) que pretendía extraer información de las redes sociales (entre otras, Facebook, Twitter y LinkedIn) para determinar la solvencia de un determinado individuo. Dicho plan no tardó en enfurecer a políticos, comentaristas y ciudadanos alemanes, que lo vieron como un ataque frontal a la privacidad de las personas. En concreto, la entonces Ministra de Justicia, Sabine Leutheusser-Schnarrenberger, exigió que "*SCHUFA y otras agencias de crédito revelen sus intenciones completas de usar los datos de Facebook para verificar la solvencia*" en una entrevista con Spiegel Online y, asimismo, Ilse Aigner, Ministra de Protección al Consumidor, manifestó al diario Münchner Merkur que "*SCHUFA no puede convertirse en el Gran Hermano del mundo empresarial*".[372] No obstante, SCHUFA aseguró que únicamente miraría los datos que estuvieran disponibles públicamente lo cual, en mi opinión, no deja de resultar polémico.

[371] Véase Die Spiegel, 2012.
[372] Die Spiegel, 2012.

Por otro lado, Parcoursup es una plataforma de admisión virtual empleada en el ámbito de la enseñanza superior francesa, operativa desde el 15 de enero del 2018 tras sustituir al sistema APB (Admission Post-bac), que permite a los alumnos del último curso escolar elegir entre más de trece mil formaciones, dependiendo de los conocimientos y competencias necesarias en esa formación, las salidas laborales y la tasa de éxito en los exámenes. [373] Dicho sistema, no obstante, emplea IA y, en concreto, el algoritmo denominado Gale-Shapley, por lo que muchos sindicatos de estudiantes se erigieron en contra de la aplicación del mismo, alegando falta de transparencia.

Tras ello, en mayo de 2018, el Ministerio de Educación Nacional y Educación Superior e Investigación francés decidió publicar el código fuente de la plataforma[374] y, asimismo, el Presidente de la República, Emmanuel Macron, durante la presentación de la Estrategia Nacional sobre IA, también expresó el deseo de que el Estado publicara los algoritmos que utiliza, incluido el de Parcoursup. [375] Los sindicatos de estudiantes, no obstante, consideraron que la publicación del código fuente de la plataforma no resultaba suficiente, por lo que iniciaron procedimientos ante varios Tribunales Administrativos para solicitar la publicación de los algoritmos locales utilizados por las universidades para determinar qué criterios se tenían en cuenta y verificar si el empleo de los mismos podía conllevar discriminación (en concreto, la sospecha era que el origen del candidato se utilizaba con fines discriminatorios, lo que llevaba a las universidades a favorecer a los candidatos de las escuelas secundarias parisinas en detrimento de los candidatos de las escuelas secundarias en los suburbios de París, con independencia de sus resultados académicos respectivos).

En relación con ello, lo que alegaban el Ministerio y la Conferencia de Presidentes de Universidades para mantener el contenido del algoritmo oculto era "el secreto de deliberación", siendo que consideraban que los criterios de clasificación de los candidatos eran un asunto que correspondía conocer únicamente a los miembros del jurado soberano. No obstante, el Defensor del Pueblo francés recomendó que se hicieran públicos los criterios precisos que empleaban las universidades para elegir a sus candidatos, al entender

[373] Véase Écoles françaises. Espagne-Portugal EFEP., 2018.
[374] Certes, 2018.
[375] 20 Minutes, 2018.

que la publicación de tal información no afectaba a los principios de la soberanía del jurado y al secreto de sus deliberaciones.[376]

El primer Tribunal en resolver, al respecto, el 4 de febrero del 2019, fue el Tribunal Administrativo de Guadalupe, que finalmente obligó a la Universidad de las Antillas a publicar los algoritmos utilizados para seleccionar a sus candidatos a través de la plataforma Parcoursup. No obstante, tal centro universitario, al advertir que la decisión del Tribunal difería de la de la autoridad administrativa independiente denominada "*Commission d'accès aux documents administratifs*" (CADA), que había considerado -en el mismo caso- que no resultaba necesario publicar los algoritmos locales empleados por las distintas universidades, acudió al Consejo de Estado francés para que aclarara cuál el marco legal aplicable.[377]

Y, por su parte, el Consejo de Estado francés, el 12 de junio del 2019, publicó una resolución en que determinó que a pesar del derecho reconocido por la Ley de 8 de marzo de 2018 a los candidatos registrados en la plataforma Parcoursup para conocer los criterios de análisis de su candidatura por parte de los establecimientos de educación superior y los motivos de la las decisiones tomadas sobre ellos, tales informaciones no estaban abiertas a los sindicatos de estudiantes, por lo que la Universidad de las Antillas pudo negarse legalmente a comunicar a un sindicato los datos relacionados con el procesamiento informático de sus aplicaciones. No obstante, a partir del Decreto de 26 de marzo de 2019, posterior a la disputa, los establecimientos de educación superior sí deben publicar los criterios generales utilizados en su procedimiento de selección.

En mi opinión, el sector público es uno de los que pueden verse más beneficiados (por el impacto a gran escala que todos sus éxitos alcanzan) por uso de la IA, puesto que, en gran medida, tal tecnología puede contribuir a hacer más eficiente (especialmente en el ámbito económico-temporal) la toma de decisiones de las en muchos casos obsoletas Administraciones Públicas, con una consiguiente y evidente mejora del servicio a los ciudadanos. No obstante, como ya he dicho, es importante poner de relieve que el principio de transparencia debe cobrar especial relevancia en el caso de los algoritmos empleados en

[376] Stromboni, 2019.
[377] Haas & Poujo, 2019.

dicho sector, ya que, además de ser inevitables, como he apuntado con anterioridad, estos deben actuar siempre en interés colectivo y suelen ser utilizados para implementar normas legales (cálculo de impuestos, por ejemplo). En el ámbito judicial, en concreto, la falta de transparencia de los algoritmos empleados para tomar decisiones que afectan a un individuo, por ejemplo, podría resultar constitutiva de una vulneración del derecho a la tutela judicial efectiva y del derecho de defensa reconocidos en los artículos 8, 10 y 11 de la Declaración Universal de Derechos Humanos y en el artículo 47 y 48 de la Carta de Derechos Fundamentales de la Unión Europea, lo cual es muy grave.

En octubre de 2019, el Joint Council for the Welfare of Immigrants (JCWI), un grupo de defensa de los derechos de los inmigrantes de Reino Unido, con el apoyo de la ONG Foxglove, anunció su voluntad de llevar a los Tribunales al Ministerio del Interior británico, con el objetivo de esclarecer qué había detrás del algoritmo empleado por este para filtrar y decidir sobre las solicitudes de visados de trabajo, ante la sospecha de la existencia de sesgos raciales en el mismo. Dicho Ministerio, no obstante, ha mantenido en todo momento que tal algoritmo se emplea únicamente para asignar y clasificar las solicitudes de forma más eficiente, pero no decide sobre el fondo de las mismas, puesto que la decisión final recae sobre trabajadores sociales (humanos, no máquinas).[378]

En Austria, la Agencia Nacional de Empleo utiliza un algoritmo que asigna a los solicitantes una puntuación relativa a sus posibilidades de hallar oportunidades en el mercado laboral, lo que que hace que estos queden clasificados en tres grupos, cada uno de los cuales tiene asignada una formación, unos beneficios y un grado de apoyo distintos. Uno de los criterios que el algoritmo tiene en cuenta para establecer la puntuación y la consiguiente clasificación, es el del sexo, lo que puede conducir a la mujer a obtener una peor valoración, razón por la que el sistema ha sido duramente criticado, habida cuenta de sus tintes discriminatorios. No obstante, desde la Agencia defienden que sus trabajadores humanos tienen la posibilidad de anular la categorización del algoritmo en caso de observar que este arroja resultados sexistas.[379]

[378] Véase McDonald, 2019.
[379] Véase Of Men & Machines, 2019.

En 2014, Polonia introdujo asimismo un sistema similar de recopilación de datos y perfiles para la clasificación de ciudadanos desempleados, lo que conllevó que la Fundación Panoptykon publicara en 2015 un informe denominado "*Profiling the unemployed in Poland: social and political implications of algorithmic decision making*"[380] criticando su impacto negativo sobre los derechos fundamentales de los ciudadanos, en concreto, el derecho a la intimidad, a la protección de datos, a la igualdad y a la no discriminación, entre otros, y advirtiendo de que, a pesar de que sus trabajadores humanos tenían la opción de anular la decisión realizada por el algoritmo, la mayoría de las veces, estos confiaban en la puntuación y clasificación realizada por este.

En los Países Bajos, en el mismo año, el Ministerio de Asuntos Sociales lanzó el programa "*Systeem Risico Indicatie*" (SyRI) creado para, a través de 17 categorías de datos gubernamentales, identificar mediante IA a las personas consideradas de alto riesgo de cometer fraude contra Hacienda y la Seguridad Social y así establecer medidas de prevención. Algunos municipios holandeses (entre ellos, Rotterdam) aplicaron tal herramienta, aunque dirigida de forma exclusiva a determinados barrios que contaban con un elevado número de residentes inmigrantes con bajos niveles de ingresos (entre ellos, Hillesluis y Bloemhof).

Tras ello, a pesar de que el gobierno holandés aseguró que SyRI resultaba una herramienta necesaria y legítima para la lucha contra el fraude, ciertos grupos de la sociedad civil empezaron a tener sospechas de que su uso podía resultar discriminatorio por razón del origen y decidieron iniciar una investigación. En 2018 se interpuso una demanda por, entre otros, la federación sindical FNV, el Consejo Nacional de Consumidores, el Comité Jurídico Holandés para los Derechos Humanos y la Plataforma para la Protección de los Derechos Civiles, en un tribunal de distrito de La Haya, habiendo contado con el apoyo del Relator especial de Derechos Humanos de las Naciones Unidas Philip Alston, quien presentó un informe, *amicus curiae*, poniendo de manifiesto que el sistema planteaba "*amenazas potenciales significativas para los derechos humanos, en particular para los más pobres de la sociedad*".[381]

[380] Niklas, Sztandar-Sztanderska & Szymielewicz, 2015.
[381] Alston, 2019, pág. 12.

El 5 de febrero del 2020, el mencionado tribunal estimó la demanda interpuesta y falló que, si bien ciertamente resultaba legítima la utilización de la tecnología por parte del Gobierno para combatir el fraude, el sistema SyRI era demasiado invasivo y podía implicar la vulneración del derecho a la privacidad de los ciudadanos, entendiendo que no se ajustaba a los principios de transparencia (puesto que no se pusieron a disposición del público -ni del tribunal- ni el modelo ni los indicadores de riesgo, ni los datos exactos utilizados), y minimización de recopilación de datos -proporcionalidad- establecidos en el RGPD, así como el principio a la igualdad y a la no discriminación previstos en el Convenio Europeo de Derechos Humanos (y en la Carta de Derechos Fundamentales de la UE), al establecer conexión entre el origen inmingrante y el bajo nivel de ingresos con un mayor riesgo de fraude. Según Christiaan van Veen, director del Proyecto de Estado de Bienestar Digital y Derechos Humanos de la Facultad de Derecho de la Universidad de Nueva York, el fallo marcó un punto de partida que excede de las fronteras holandesas.[382] [383]

En España, el 20 de junio del 2019 la Fundación Civio presentó un recurso contencioso-administrativo ante la negativa del Consejo de Transparencia de obligar a hacer público el código fuente del programa BOSCO, que decide quién resulta o no beneficiario del bono social eléctrico. En la demanda se dispuso: "*Cuando el código fuente de un programa informático es ley, porque mediante su ejecución se generan derechos y obligaciones, el ciudadano tiene tanto derecho a inspeccionar su funcionamiento como lo tiene con respecto a cualquier otra norma jurídica*", y añade "*Impedir el acceso al código fuente de los programas implica la imposibilidad de que el ciudadano pueda verificar si una herramienta a través de la cual se le aplica el derecho se halla acorde con la ley o incumple la misma*".[384]

Y, en relación con ello, Andrés Boix Palop, Profesor titular de Derecho administrativo de la Universidad de Valencia, en la misma línea argumental que la mencionada demanda, hace una interesante reflexión sobre el valor que debería otorgarse a los algoritmos en el ámbito del sector público, en su opinión, el de reglamentos. Y es que argumenta que los algoritmos utilizados por las Administraciones Públicas para la toma de decisiones deben

[382] Véase van Veen, 2020.
[383] Véase Fernández, 2020.
[384] Civio, 2019.

tener tal consideración, habida cuenta de que cumplen una función material equivalente a la de las normas jurídicas, al reglar y predeterminar la actuación de los poderes públicos. Así, el mencionado autor entiende que tales herramientas de IA han de someterse a estrictos procedimientos de elaboración y aprobación, su contenido ha de estar debidamente publicado y deben contar con mecanismos de recurso directo e indirecto frente a ellos, lo cual, no me parece en absoluto descabellado. [385]

En relación con ello, tal y como pone de manifiesto Mª Jesús González-Espejo, Directora del Instituto de Innovación Legal: "*Como muestran todos estos diferentes casos, está claro que mientras no exista una autoridad responsable de realizar un control previo de la legalidad de la IA, dicho control parece que tendrá que seguir recayendo en la sociedad civil y cuando ésta no tenga la capacidad de paralizar el uso de herramientas sobre las que se tenga la presunción de que no están respetando de alguna forma nuestros derechos fundamentes, entonces resultará necesario recurrir a un juez para exigir que deje de usarse*."[386]

Y es que es importante traer a colación el denominado Derecho a una buena Administración reconocido en el artículo 41 de la Carta de Derechos Fundamentales de la UE, que incluye de forma específica el derecho a la igualdad, a la transparencia, a la seguridad y a la responsabilidad, y dispone (con subrayado propio):

"*1. Toda persona tiene derecho a que las instituciones y órganos de la Unión traten sus asuntos imparcial y equitativamente y dentro de un plazo razonable.*

2. Este derecho incluye en particular:

-el derecho de toda persona a ser oída antes de que se tome en contra suya una medida individual que le afecte desfavorablemente,
-el derecho de toda persona a acceder al expediente que le afecte, dentro del respeto de los intereses legítimos de la confidencialidad y del secreto profesional y comercial,
-la obligación que incumbe a la administración de motivar sus decisiones.

[385] Boix, 2020, págs. 223-270.
[386] González-Espejo, 2020.

3. Toda persona tiene derecho a la reparación por la Comunidad de los daños causados por sus instituciones o sus agentes en el ejercicio de sus funciones, de conformidad con los principios generales comunes a los Derechos de los Estados miembros.

4. Toda persona podrá dirigirse a las instituciones de la Unión en una de las lenguas de los Tratados y deberá recibir una contestación en esa misma lengua."

C) Principio de equidad, igualdad, no discriminación del ser humano e inclusión

Uno de los problemas que más ríos de tinta ha hecho correr en el ámbito de la IA es el de la posible existencia de sesgos en los algoritmos, lo que, sin duda, podría vulnerar el derecho a la no discriminación reconocido en el artículo 2 de la Declaración Universal de Derechos Humanos, el artículo 14 del Convenio Europeo de Derechos Humanos, el artículo 21 de la Carta de Derechos Fundamentales de la Unión Europea y el artículo 14 de la Constitución Española. Y es que si bien es evidente que los riesgos de discriminación deben ser prevenidos y eliminados en cualquier circunstancia, debe prestarse especial atención a aquellos grupos cuyos derechos tienen mayor riesgo de verse afectados por la IA.

Tal y como ya se ha avanzado a lo largo de la presente tesis doctoral, la humanidad se enfrenta a una nueva era y se halla, sin duda, en un punto de inflexión clave y crucial que puede implicar un sustancial cambio positivo para nuestras vidas, pero ello solo ocurrirá en el caso de que se tomen decisiones orientadas a obtener un beneficio colectivo. Y es que a lo largo de la historia se han ido arrastrando comportamientos y creencias muy perjudiciales para ciertos grupos o minorías de la población, y ello, mediante las nuevas tecnologías (y en concreto la IA), puede resultar finalmente nuetralizado o, al menos, muy minimizado. No obstante, si no se presta la debida atención y no se hace un esfuerzo a nivel nacional e internacional, resulta muy posible que esos potenciales avances se conviertan en meras perpetuaciones de conductas y convicciones pasadas y, en vez lograr un progreso que mejore nuestras sociedades, hagan que permanezcan para siempre entre nosotros los fantasmas del pasado.

En el ámbito de la discriminación por razón de género, resulta especialmente interesante el informe publicado en mayo del 2019 por la UNESCO, en colaboración con el Gobierno de

Alemania y The EQUALS Skills Coalition -una alianza del sector público y privado que fomenta la participación de mujeres y niñas en el ámbito de la tecnología- titulado "*I'd blush if I could-Closing gender divides in digital skills through education*".[387] El título del mismo hace referencia a la respuesta estándar que la voz femenina de la asistente digital de Apple, Siri, da cuando recibe insultos de los usuarios. Si bien el estudio está centrado principalmente en poner de relieve el papel de la educación para conseguir un mayor nivel de igualdad entre hombres y mujeres en esta nueva era digital, una de las cuestiones que menciona es el hecho de que, además de Siri, múltiples asistentes virtuales que emplean IA tienen voz de mujer y adoptan papeles sumisos, lo que supone un sesgo de género en el sistema que no viene más que a continuar con una concepción retrógrada y contraria al derecho a la igualdad que tantos perjuicios ha ocasionado a lo largo de los tiempos. Entre otras, puede hacerse referencia a Alexa (asistente virtual de Amazon), Google Home (asistente virtual de Google), Kiri (asistente virtual de Vodafone) o Sophie (asistente virtual de Air New Zealand). Asimismo, al menos en España, la voz que por defecto se escucha cuando se emplean los navegadores Google Maps o el sistema Tom Tom, es femenina, lo que no viene más que a perpetuar el rol de "copiloto" o asistente de la mujer, frente al papel de "conductor principal" del hombre.

Tal y como ya se ha avanzado con anterioridad, no obstante, los algoritmos, *per se*, son neutros. Y es que, está claro que estos toman forma según el diseño de quien los crea, por lo que el foco debe ponerse principalmente en ese proceso de creación previo en que se sientan las bases del posterior contenido del sistema de IA. En relación con ello, de forma contínua me he topado con ingenieros que reclaman una mayor y más fluida y estrecha colaboración entre los profesionales técnicos y los jurídicos, habida cuenta de las infinitas dudas legales que les surgen a la hora de proyectar y programar sistemas de IA, principalmente en relación a las cuestiones de la posible vulneración de derechos fundamentales. Y es que, por parte de los técnicos se reclaman a los juristas definiciones y pautas más claras y concretas sobre lo que implican y significan la equidad, la igualdad y la no discriminación, entre otros. Es cuestión comúnmente aceptada el hecho de que los creadores de sistemas de IA, con unas directrices claras y específicas, pueden diseñar

[387] Véase UNESCO, 2019.

algoritmos y programas que carezcan (o reduzcan a la mínima expresión) los sesgos, si bien para ello resulta necesario un trabajo colaborativo y multidisciplinar.

Un avance en tal sentido se puede hallar en el Anexo de Recomendaciones a la Comisión y al Consejo de la UE, de la Resolución del Parlamento Europeo de 16 de febrero del 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho Civil sobre robótica, que establece un "Código de conducta ética para los ingenieros en robótica" y un "Código deontológico para los comités de ética de la investigación", con pautas claras, y un apartado denominado "Licencia para los diseñadores", que empieza exigiendo "*Los diseñadores deberán tener en cuenta los valores europeos de dignidad, autonomía y autodeterminación, libertad y justicia, antes, durante y después del proceso de concepción, desarrollo y de aplicación de esas tecnologías, incluida la necesidad de no perjudicar, herir, engañar o explorar a los usuarios (vulnerables)*."

En relación con ello, tal y como se expone en el artículo publicado por los investigadores de la Universidad de Stanford (California, EEUU) Sam Corbett-Davies y Sharad Goel titulado "*The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*" (con surbayado propio): "*En los últimos años, la comunidad científica ha propuesto una multitud de definiciones formales y matemáticas de equidad para ayudar a los profesionales a diseñar herramientas equitativas de evaluación de riesgos. En particular, tres amplias clases de definiciones de equidad han cobrado importancia. La primera, a la que llamamos anti-clasificación, propone que los algoritmos de evaluación de riesgos no tengan en cuenta las características protegidas, tales como la raza, el género o similares, cuando realicen estimaciones. La segunda clase de definiciones exige paridad de clasificación, lo que requiere que ciertas medidas comunes de predicción del comportamiento sean equitativas en los grupos definidos por los atributos protegidos. Bajo esta definición, un algoritmo de evaluación de riesgos que predice el incumplimiento de un préstamo podría, por ejemplo, ser requerido para producir tasas falsas negativas similares tanto para los solicitantes blancos como los negros. Finalmente, la tercera definición formal de equidad, conocida como calibración, requiere que los resultados sean independientes de los atributos protegidos después de controlar el riesgo estimado. Por ejemplo, entre los solicitantes de préstamos que se estima que tienen un 10% de*

probabilidad de incumplimiento, la calibración requiere que los blancos y los negros incurran en incumplimiento en tasas similares." [388]

No obstante, tal y como pone de manifiesto el mencionado estudio, estas tres clasificaciones o categorías tienen profundas limitaciones estadísticas, resultando medidas poco eficaces para detectar algoritmos discriminatorios y, todavía más preocupante, diseñar algoritmos para satisfacer y cumplimentar tales definiciones puede generar un efecto negativo en el bienestar tanto de las minorías como de las mayorías. Y es que, por ejemplo, en contra de lo que dispone el principio de anti-clasificación, a menudo es necesario que los algoritmos de evaluación de riesgos, para ser equitativos, consideren de forma expresa las características protegidas. Así, en el ámbito judicial penal, por ejemplo, suele ser más probable que los hombres cometan un delito violento en comparación con las mujeres con antecedentes similares y, como resultado de ello, medidores de riesgo neutros en relación con el género ("*gender-neutral*") pueden sobreestimar, de forma sistematica, el nivel de riesgo de reincidencia de las mujeres, lo que a su vez podría fomentar decisiones judiciales incorrectas o desproporcionadas. Conscientes de tal riesgo, algunas jurisdicciones, tales como la de Wisconsin (EEUU), han decidido emplear herramientas de evaluación de riesgo distintas según el sexo ("*gender-specific*") con el fin de asegurar que las predicciones no están sesgadas en contra de las mujeres.

Es interesante, no obstante, en tal sentido, hacer mención del estudio publicado por de Jon Kleinberg, Sendhil Mullainathan y Manish Raghavan[389], en que se asegura y demuestra que no resulta posible diseñar algoritmos justos para tomar cierta clase de decisiones.

A la vista de todo lo expuesto, es importante, pues, el fomento de una IA inclusiva, accesible para todos por igual y que no aumente todavía más la brecha entre países ricos/pobres, minorías/mayorías que, por desgracia, existe todavía en estos tiempos.

En relación con ello, es interesante traer a colación el dato que se recoge en el Considerando O de la Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de IA y robótica: "*Considerando que el desarrollo de*

[388] Corbett-Davies & Goel, 2018, pág. 2.
[389] Véase Kleinberg, Mullainathan & Raghavan, 2017.

la IA y de la robótica debe incluir a la sociedad en su conjunto; que, no obstante, en 2017 las zonas rurales seguían excluidas en gran medida de los beneficios de la IA, ya que el 8 % de los hogares no tenían acceso a ninguna red fija y el 53 % no disponía de ninguna tecnología de «acceso de próxima generación» (VDSL, Cable Docsis 3.0 o FTTP)." Y, asimismo, también es importante poner el foco en lo dispuesto en el Considerando AE, que hace referencia a la necesidad de afrontar, en el ámbito de la UE (aunque la problemática es extensible a nivel mundial), las diferencias lingüísticas y los posibles riesgos de exclusión que acechan a aquellas lenguas minoritarias y, por ende, a los ciudadanos que las hablan: "*Considerando que la Inteligencia Artificial debe apoyar plenamente todas las lenguas europeas para ofrecer a todos los europeos las mismas oportunidades de beneficiarse de la evolución moderna de la inteligencia artificial en la sociedad de la información multilingüe europea*".

Y es que, desde luego, realizar grandes inversiones en la investigación, creación y desarrollo de sistemas de IA sin contar con los medios necesarios para que estos puedan otorgar oportunidades y beneficios a todas (o la gran mayoría) las personas y las sociedades por igual, lo único que va a provocar es un aumento todavía mayor de las desigualdades en el mundo, habida cuenta de los enormes avances que tal tecnología puede provocar, lo que, sin duda alguna, no hará más que dejar, aun más atrás, a las minorías sociales o a los países más necesitados.

En relación con ello, la denominada "alfabetización digital", entendida como la educación y la formación (inicial y contínua) en materia de nuevas tecnologías (entre las que se halla la IA) debe ser sin duda una de las prioridades de las políticas que se diseñen en el ámbito de la IA (y, de hecho, ya lo es), puesto que es la clave para fomentar una participación integradora y amplia en relación con tal tecnología.

En mi opinión, la mejor solución sería la creación de unas guías o manuales elaborados por profesionales de diversos ámbitos y avaladas por una agencia u organismo público que establecieran protocolos de actuación claros, que deberían ser o bien flexibles y comunes, para poder resultar de aplicación en los máximos ámbitos posibles. Se reputa fundamental que tales planes sean analizados y verificados por un organismo público (que bajo mi punto de vista, debería ser el mismo que el que emitiera certificaciones de calidad de los sistemas

de IA, al que se ha hecho alusión con anterioridad), a nivel nacional, europeo o internacional, puesto que la unificación de criterios, la garantía de protección de los derechos fundamentales y la seguridad jurídica resultan imprescindibles para conseguir un buen uso y aplicación de los sistemas de IA y para forjar un clima de confianza en la población.

D) Principio de robustez, solidez técnica y seguridad

Resulta evidente que, cada vez mas, los seres humanos iremos utilizando y sirviéndonos de sistemas de IA para desarrollar no solo las actividades más básicas sino también las más complejas. Ello, no obstante, y a pesar de suponer un gran avance en eficiencia, bienestar y reducción de costes, puede implicar grandes peligros. Y es que el ser humano, por naturaleza, tiende a "acomodarse". Así, si bien en ciertas ocasiones dar los primeros pasos ante algo novedoso puede resultar inquietante o difícil, lo cierto es que una vez se conoce el nuevo terreno que se pisa, tendemos a "dejarnos llevar". Piénsese, por ejemplo, en la cautela con la que se conduce las primeras veces y en lo mecánica que se convierte tal tarea cuando ya se tiene más experiencia y se ha comprobado que el vehículo responde; o, piénsese en la animadversión que existía al principio hacia los teléfonos móviles y la dependencia que hoy en día existe respecto de ellos, que han pasado a formar parte indispensable de nuestras vidas, siendo que los manjeamos de forma totalmente automática y muchas veces inconsciente.

Tal comportamiento humano, en el ámbito de la IA, resulta especialmente preocupante, y es que es muy fácil caer en el denominado fenómeno "sesgo de automatización", que implica una dependencia excesiva de dicha automatización. Y es que el hecho de confiar demasiado en un sistema automatizado hace que los humanos disminuyamos significativamente la búsqueda, la atención y el procesamiento de la información (piénsese, por ejemplo, en los viajes que realizamos en coche en la actualidad, en que ni siquiera, la mayoría de veces, pensamos o planeamos la ruta que vamos a tomar, puesto que simplemente confiamos en que los sistemas de GPS nos guiarán por los mejores caminos).

En relación con ello, según los investigadores Raja Parasuraman y Dietrich H. Manzey [390] existen, en términos generales, tres factores principales para el sesgo de automatización:

1) los humanos, si tenemos la opción de tomar decisiones, elegimos el camino que nos implique el menor esfuerzo cognitivo;
2) los humanos percibimos que los sistemas que toman de decisiones de forma automatizada son superiores a nosotros en sus capacidades y tendemos a sobreestimar su rendimiento;
3) la holgazanería social no queda reducida a los equipos humanos, sino que también aparece cuando humanos y máquinas trabajan juntos.

Así, en caso de que un humano tenga la posibilidad de acudir a un sistema de IA para sustituir una tarea que le supone un esfuerzo (principalmente intelectual) o pueda, al menos, beneficiarse de su ayuda, en la mayoría de ocasiones optará por disminuir su nivel de conciencia y atención y caerá en la tentación de "abandonarse" frente al mismo, y a pesar de que se otorgue a los humanos la opción de cuestionar y anular a los algoritmos, es muy probable que no hagan uso de tal posibilidad.

Y, tal confianza depuesta en los sistemas de IA puede implicar una automatización tal de las tareas que, en caso de fallo, los humanos que suelen emplearlos no tengan idea de cómo reaccionar. Por ello, y por el potencial daño (accidental o intencionado) que puede ocasionarse, que es muy elevado, la robustez y la solidez técnica de los sistemas son necesarias para evitar cualquier posible error o fallo, resultando necesaria para ello la implementación de mecanismos de reparación eficaces y accesibles.

En tal sentido, en la Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones, de 8 de abril del 2019, titulada "Generar confianza en la Inteligencia Artificial centrada en el ser humano" se establece que "*los sistemas de IA deben integrar mecanismos de seguridad desde el diseño para garantizar que sean verificablemente seguros en cada fase, teniendo muy presente la seguridad física y psicológica de todos los afectados*".[391] Asimismo, el Informe de la

[390] Parasuraman & Manzey, 2010.
[391] Véase Comisión Europea, 2019, pág. 5.

Comisión al Parlamento Europeo, al Consejo y al Comité Económico y Social Europeo, de 19 de febrero del 2020, titulado "Informe sobre las repercusiones en materia de seguridad y responsabilidad civil de la Inteligencia Artificial, el internet de las cosas y la robótica"[392], remarca la importancia de contar con niveles elevados de seguridad de los productos y sistemas que incorporen nuevas tecnologías digitales para garantizar la protección a los consumidores y generar confianza en tales tecnologías, y analiza y evalúa la normativa de la UE en tal materia, con el fin de determinar si se cuenta con los elementos necesarios para garantizar que las tecnologías emergentes y, en concreto, los sistemas de IA operen de forma segura en el mercado. En relación con ello, se alude a la normativa sectorial, a la Directiva 2001/95/CE del Parlamento Europeo y del Consejo, de 3 de diciembre de 2001, relativa a la seguridad general de los productos y a las legislaciones nacionales, y se identifican algunos riesgos y posibles soluciones, en concreto, entre otros:

-la conectividad, que hace que los sistemas de IA puedan sufrir ciberataques o puedan padecer terribles consecuencias en casos de pérdida de conexión de red, lo que exige la incorporación en la normativa de disposiciones expresas y específicas para mitigar tales riesgos;
-la autonomía, que puede comportar la causación, por parte de los sistemas de IA, de resultados no previstos y no deseados por sus diseñadores, fabricantes y/o usuarios, lo que hace necesaria la exigencia de un control de evaluación de riesgos con carácter previo a la comercialización y de un mecanismo de control posterior, cuando el sistema ya se halle en funcionamiento. En relación con ello, no obstante, la Comisión plantea que los actos normativos de la UE contemplen requisitos específicos de supervisión humana -que sirvan como salvaguarda-, durante el diseño y todo el ciclo de vida de los productos y sistemas de IA;
-riesgos para la salud mental de los usuarios, en aquellos casos de interacción con robots y sistemas de IA humanoide, reputándose por ello necesaria la inclusión en la normativa de la UE de obligaciones específicas consistentes en la valoración expresa del daño mental que los productos pueden causar a los usuarios (especialmente, a los más vulnerables);
-la dependencia de los datos, lo que hace que resulte imprescindible que estos sean exactos y correctos para garantizar el éxito de las decisiones de los sistemas de IA, por lo que la

[392] Véase Comisión Europea, 2020.

Comisión plantea que la normativa de la UE en materia de seguridad de los productos incluya requisitos específicos, en la fase de diseño, relacionados con el riesgo para la seguridad derivado del uso de datos erróneos, así como mecanismos para garantizar la calidad de los datos cuando se usan productos y sistemas de IA; y
-la incorporación de sistemas informáticos en las herramientas de IA que pueden poner en jaque a su seguridad, lo que hace necesaria la imposición de obligaciones adicionales a los fabricantes para evitar que la introducción de programas informáticos afecte a la seguridad de los productos de IA durante su vida útil.

Tal y como afirma el investigador Seán Ö hÉigeartaith, del "Centre for the Study of Existential Risk" (CSER) (Cambridge, Reino Unido), respecto del riesgo de los sistemas de IA "*ya sea por un uso deliberado o por circunstancias inesperadas fuera de control, podría ser superior al de cualquier tecnología de la historia humana. Si es plausible que se alcance ese nivel tecnológico durante el presente siglo, en las décadas previas será necesario un grado elevado de investigación y de planificación, tanto en el diseño técnico de dichos sistemas como en las estructuras de gobernanza alrededor de su desarrollo para conseguir una transición deseable*."[393]

Y es que uno de los principales focos de riesgo de los sistemas de IA, por su alta vulnerabilidad (incluso en los más potentes) es la inseguridad de los mismos, habida cuenta de que, siendo que la mayoría operan con *software* y mantienen conexiones con la red de Internet, estos no están libres de sufrir ciberataques, principalmente por la información tan sensible y "suculenta" que pueden llegar a manejar (datos personales relativos, incluso, a los aspectos más íntimos de la vida de las personas, a saber, la salud; sus eventuales asuntos con la justicia; su economía, etc; e industriales, incluyendo informaciones críticas que deben quedar dentro del secreto empresarial).

Es una circunstancia sobradamente conocida que los ataques informáticos o ciberataques son contínuos y estos pueden responder a múltiples intereses (económicos, políticos, criminales, etc), tanto individuales como grupales, estando cada vez más profesionalizados, con empleo de herramientas más sofisticadas y difíciles de detectar, neutralizar y, sobre

[393] Ó hÉigeartaigh, 2020.

todo, de rastrear. En relación con ello, además, es importante poner de relieve que no solo los sistemas de IA pueden verse amenazados por ciberataques, sino que la propia IA puede ser una amenaza en sí misma para la ciberseguridad, y en tal sentido en la Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de IA y robótica, se propone que la Agencia de Seguridad de las Redes y de la Información de la Unión Europea (ENISA) prepare un plan de acción para la ciberseguridad en el ámbito de la IA, que deberá evaluar y abordar sus amenazas y deficiencias específicas.

De acuerdo con ello, en el informe publicado en septiembre del 2019 por el "*European Parliamentary Research Service*" del Parlamento Europeo, se establece: "*En la práctica, deben tenerse en cuenta todas las vulnerabilidades al crear algoritmos. Ello requiere probar sistemas de IA para comprender y mitigar los riesgos de ataques cibernéticos y piratería informática. Los desarrolladores de IA deben implementar procesos capaces de evaluar los riesgos de seguridad asociados en caso de que alguien use el sistema de IA que están construyendo con fines dañinos. Por ejemplo, si el sistema se ve comprometido, debería ser posible que el control humano tome el control y lo anule. Para abordar esta importante cuestión, la UE aplica un enfoque doble: primero, fomentar la cooperación entre la comunidad de IA y la comunidad de seguridad, y segundo, reflexionar sobre cómo modificar el marco legal que rige las responsabilidades en la UE, y pasar de un régimen de responsabilidad basado en la conducta humana a un régimen de responsabilidad más basado en la conducta maquinal*."[394]

Para tratar de mitigar tales efectos, no obstante, en el ámbito nacional e internacional existen múlitples regulaciones, no solo de naturaleza civil sino también, desde luego, penal.

Entre otras, resulta interesante destacar, a nivel europeo, la Directiva 2013/40/UE del Parlamento Europeo y del Consejo, de 12 de agosto del 2013, relativa a los ataques contra los sistemas de información, por la que se sustituye la Decisión marco 2005/222/JAI del Consejo; la Decisión del Consejo relativa a medidas restrictivas contra los ciberataques que amenacen a la Unión o a sus Estados miembros, publicada el 16 de mayo del 2019; y

[394] Madiega, 2019, pág. 4.

el informe publicado el 3 de diciembre del 2019 por el Consejo Europeo titulado "Conclusiones del Consejo sobre la importancia de la tecnología 5G para la economía europea y la necesidad de mitigar los riesgos para la seguridad relacionados con tal tecnología"[395].

Particularmente interesante, en tal sentido, es hacer mención a lo dispuesto en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley De Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, que en su Considerando 28 establece: "(…) *Los sistemas de IA pueden tener efectos adversos para la salud y la seguridad de las personas, en particular cuando funcionan como componentes de productos. En consonancia con los objetivos de la legislación de armonización de la Unión de facilitar la libre circulación de productos en el mercado interior y velar por que solo lleguen al mercado aquellos productos que sean seguros y conformes, es importante prevenir y reducir debidamente los riesgos de seguridad que pueda generar un producto en su conjunto debido a sus componentes digitales, entre los que pueden figurar los sistemas de IA*. (…)". Y, a lo largo de la Propuesta, se establecen medidas de seguridad, especialmente para los sistemas calificados de "alto riesgo", con el fin de disminuir los efectos adversos que estos podrían generar en los seres humanos en caso de fallo.

Y, a nivel nacional, debe hacerse alusión, entre otras, a la Ley 34/2002, de 11 de julio, de Servicios de la Sociedad de la Información y de Comercio Electrónico; en parte, a la Orden PRE/2740/2007, de 19 de septiembre, por la que se aprueba el Reglamento de Evaluación y Certificación de la Seguridad de las Tecnologías de la Información; a la Ley 25/2007, de 18 de octubre, de Conservación de Datos relativos a las Comunicaciones Electrónicas y a las Redes Públicas de Comunicaciones (que traspuso la Directiva 2006/24/CE, del Parlamento Europeo y del Consejo, de 15 de marzo del 2006); especialmente a la Ley 8/2011, de 28 de abril, por la que se establecen medidas para la Protección de las Infraestructuras Críticas; al Real Decreto 704/2011, de 20 de mayo, por el que se aprueba el Reglamento de Protección de las Infraestructuras Críticas; a la Ley Orgánica 4/2015, de 30 de marzo, de Protección de la Seguridad Ciudadana; a la Ley 36/2015, de 28 de

[395] Véase Consejo de la Unión Europea, 2019.

septiembre, de Seguridad Nacional; y a algunos preceptos del Código Penal (entre otros, los artículos 197 bis, 197 ter, 197 quater, 197 quinquies).

Una posible solución a la problemática planteada, desde luego, podría ser la creación de una comisión/división técnica ubicada dentro de la Agencia (nacional, europea o incluso internacional) de certificación de sistemas de IA a la que y he hecho referencia en más de una ocasión a lo largo de la presente tesis doctoral, para que analice, revise y evalúe el diseño, la fabricación y el funcionamiento de estos con carácter previo a autorizar su uso, comercialización y/o distribución en aras de garantizar su seguridad y robustez técnica. Y es que ello permitiría, sin duda, prevenir o limitar los eventuales fallos y vulnerabilidad de los sistemas, lo cual resulta fundamental para fomentar la anhelada confianza de los humanos en la IA.

En relación con ello, un buen punto de partida podría ser el Anexo de Recomendaciones a la Comisión y al Consejo de la UE, de la Resolución del Parlamento Europeo de 16 de febrero del 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho Civil sobre robótica, que tal y como ya se ha avanzado, establece un "Código de conducta ética para los ingenieros en robótica" y pone de manifiesto la importancia de la seguridad, disponiendo: "*Los diseñadores de robots han de tener en cuenta y respetar la integridad física, la seguridad, la salud y los derechos de las personas. Un ingeniero en robótica debe preservar el bienestar sin dejar de respetar los derechos humanos, y divulgar con prontitud los factores susceptibles de poner en peligro a la población o al medio ambiente*."

Asimismo es importante hacer referencia, en el ámbito europeo, al Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a la Agencia de la Unión Europea para la Ciberseguridad y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) nº526/2013 ("Reglamento sobre la Ciberseguridad"), que establece un sistema de esquemas de certificación a escala de la UE y una Agencia de la UE para la Ciberseguridad (para sustituir a la Agencia de Seguridad de las Redes y de la Información de la Unión Europea (ENISA)).

Y, asimismo, en el ámbito español, resulta pertinente hacer alusión a la Orden PRA/33/2018, de 22 de enero, por la que se publica el Acuerdo del Consejo de Seguridad Nacional, que regula el Consejo Nacional de Ciberseguridad, y al Real Decreto 421/2004, de 12 de marzo, que regula el Centro Criptológico Nacional.

Es importante poner de manifiesto que, justamente con la finalidad de crear sistemas de IA en colaboración con los reguladores, existen ya campos de prueba reglamentarios o *sandboxes*, que son entornos controlados que tienen como finalidad la introducción y prueba de ideas innovadoras con el fin de evaluar en un entorno real el uso seguro y eficaz de los sistemas de IA. Y ello se prevé asimismo en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, que en el Título Quinto regula los denominados "Espacios controlados de pruebas para la IA" que, según dispone el artículo 53: "*proporcionarán un entorno controlado que facilite el desarrollo, la prueba y la validación de sistemas innovadores de IA durante un periodo limitado antes de su introducción en el mercado o su puesta en servicio, en virtud de un plan específico*."

Y es que, como se ha visto, está claro que un sistema de IA puede contener fallos que provoquen resultados no previstos o no deseados por sus creadores y estos, sin duda, pueden ocasionar daños irreversibles en los usuarios, por lo que resulta fundamental imponer unos estándares de calidad claros y exigentes, así como crear un organismo independiente que controle el cumplimiento de los mismos para garantizar la seguridad.

Así lo recoge el "Libro Blanco sobre IA-Un enfoque europeo orientado a la excelencia y la confianza", publicado por la Comisión Europea el 19 de febrero del 2020, cuyo apartado 5, titulado "Un Ecosistema de Confianza: El Marco Regulador de la IA", que contiene el subapartado "Riesgos para la seguridad y el funcionamiento eficaz del régimen de responsabilidad civil" dispone: "*Las tecnologías de IA pueden presentar nuevos riesgos de seguridad para los usuarios cuando estén integradas en productos y servicios. Por ejemplo, como resultado de un defecto en la tecnología de reconocimiento de objetos, un vehículo autónomo puede detectar erróneamente un objeto en la carretera y causar un accidente que provoque heridos y daños materiales. Como sucede con los riesgos para los derechos fundamentales, estos riesgos pueden proceder de defectos en el diseño de la*

tecnología de IA, estar relacionados con problemas de disponibilidad o calidad de los datos, u otros derivados del aprendizaje de las máquinas. Aunque algunos de estos riegos no se limitan a los productos o servicios que dependen de la IA, el uso de esta última puede aumentar o agravar los riesgos."

Y es que, tal y como ya se ha ido apuntando, los sistemas de IA, sobre todo de *Deep Learning*, tienden a ser cada vez más autónomos, por lo que, salvo que se establezcan prohibiciones al respecto, se antoja imposible evitar que estos actúen de forma discrecional e imprevista para los usuarios, e incluso para sus diseñadores. Y ello, no obstante, no podría reputarse un "fallo técnico", sino una circunstancia inherente al propio sistema que, no obstante, puede implicar la toma de decisiones muy perjudiciales para el ser humano. Así pues, lo que tenemos que decidir los humanos, como sociedad, es qué grado de imprevisibilidad de tales sistemas de IA estamos dispuestos a aceptar, y eso, en mi opinión, solo puede resultar decidido de forma democrática y resultar plasmado por vía legislativa o reglamentaria, a nivel nacional o internacional.

Como consecuencia de lo expuesto, y debiendo ser absolutamente conscientes del enorme riesgo que exige (por la concentración de poder que se otorga a las máquinas), deviene fundamental que las herramientas de IA tengan un diseño y un funcionamiento robusto, que cumpla con todos los estándares éticos y que cuente con una solidez técnica que evite mayores perjuicios, que en muchos casos pueden ser irreversibles.

Finalmente, y a modo de apunte, es importante hacer mención a la posibilidad de que los sistemas de IA se empleen de forma malintencionada por parte de sus diseñadores o sus usuarios, con fines distintos a los legalmente permitidos, lo cual podría sin duda constituir una amenaza para la seguridad y el respeto a los derechos fundamentales de los ciudadanos, habida cuenta del enorme potencial nocivo que ello alberga. En tal sentido, la Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de IA y robótica, anima a que "*la investigación en el campo de la IA también se centre en la detección de casos en los que la IA y la robótica hayan sido manipulados de forma accidental o malintencionada*".[396] Y es que, sin duda, es un foco de

[396] Punto 11.

peligro a tener muy en cuenta, puesto que es un fenómeno creciente y las consecuencias de no prestarle atención podrían ser fatales.

Lo que, en cualquier caso, resulta evidente a la vista de lo expuesto es la necesidad de establecer la obligación de realizar auditorías independientes internas y externas de forma contínua a los sistemas de IA, especialmente a aquellos cuyo uso afecta los derechos fundamentales. En concreto, Silvia Díaz Alabart, Catedrática de Derecho Civil de la Universidad Complutense de Madrid, propone que los robots, desde que comienzan su vida útil, sean sometidos a un seguimiento de su estado técnico, con revisiones periódicas realizadas por la Administración pública[397], como en el caso de la Inspección Técnica de Vehículos (ITV) que pasan los vehículos en nuestro país.

Y ello, asimismo, se recoge como medida en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, que para los sistemas calificados de "alto riesgo" prevé un plan de revisión continuada a lo largo de toda la vida de estos. Así, el artículo 9 del mencionado texto legal dispone:

"*1. Se establecerá, implantará, documentará y mantendrá un sistema de gestión de riesgos asociado a los sistemas de IA de alto riesgo.*
2. El sistema de gestión de riesgos consistirá en un proceso iterativo continuo que se llevará a cabo durante todo el ciclo de vida de un sistema de IA de alto riesgo, el cual requerirá actualizaciones sistemáticas periódicas. (…)"

Por su parte, el artículo 44, respecto de los certificados de conformidad con que deben contar los sistemas para operar en la UE, se establece en su apartado 2 su temporalidad, con el fin de que se lleven a cabo revisiones periódicas:

"*Los certificados serán válidos para el período que indican, que no excederá de cinco años. A solicitud del proveedor, la validez de un certificado podrá prorrogarse por*

[397] Rogel, Lacruz, Mozo & Díaz, 2018, pág. 111.

períodos renovables no superiores a cinco años, sobre la base de una nueva evaluación con arreglo a los procedimientos de evaluación de la conformidad aplicables."

Y, asimismo, el Título VIII, que se titula "*Seguimiento posterior a la comercialización, intercambio de información y vigilancia de mercado*", prevé en su Capítulo I la realización de un seguimiento posterior a la comercialización de los sistemas de IA por parte de los proveedores y, en concreto, un plan de seguimiento posterior a la comercialización para sistemas de IA de alto riesgo, en los términos previstos en el artículo 61; y en su Capítulo II una obligación a los proveedores de notificar a las autoridades de vigilancia todos aquellos incidentes graves y fallos de funcionamiento de los sistemas de IA de alto riesgo introducidos en el mercado de la UE que constituyan un incumplimiento de las obligaciones de Derecho de la UE destinadas a proteger los derechos fundamentales.

En cualquier caso, es fundamental poner de manifiesto la necesidad de cooperación internacional en materia de exigencias de seguridad de los sistemas de IA, puesto que la globalización implica una expansión de los eventuales riesgos que pueden surgir de la falta de cuidado en tal sentido, lo cual derivaría en daños globales muy difíciles de mitigar.

E) Principio de responsabilidad

El principio de responsabilidad hace referencia a dos cuestiones: por un lado, a la necesidad de que se lleve a cabo un diseño, desarrollo y uso de la IA responsable, es decir, consciente, prudente, respetuoso con los derechos humanos, atento a las implicaciones éticas que puede conllevar y tendente a garantizar la seguridad; y, por otro lado, a la existencia de mecanismos de respuesta y rendición de cuentas en relación con los posibles daños que tal tecnología puede ocasionar.

Y es que una de las preguntas más frecuentes y complejas que han surgido, de forma justificada, en torno a la IA, es: ¿quién responde de los posibles efectos causados por tal tecnología?. Pues bien, no resulta fácil la respuesta.

Tal debate empezó a cobrar especial importancia tras los accidentes de vehículos autónomos ocurridos en Estados Unidos los días 19 y 23 de marzo del 2018 en EEUU.

Y es que cierto es que la tecnología de tales vehículos tiene un potencial enorme para mejorar la seguridad vial, habida cuenta de que se calcula que el 90% de los accidentes de tráfico se deben a errores humanos.[398] No obstante, en caso de que existan fallos, es importante garantizar que los daños que puedan ocasionarse sean imputables a un responsable y se reparen de forma eficiente.

Por un lado, en el caso del accidente del 19 de marzo ocurrido en Arizona (EEUU), el vehículo autónomo -fabricado por la compañía Volvo, al servicio de UBER, con supervisión de uno de sus conductores- arrolló a una ciclista, a la que causó la muerte.

En relación con ello, es interesante apuntar que ya el 3 de diciembre del 2016 Volvo publicó en su blog "Tecvolución" un informe titulado ¿De quién es la culpa del accidente cuando hay un coche autónomo de por medio? en que indicó claramente cuál era la postura oficial de la compañía acerca de la eventual responsabilidad por daños causados por un vehículo autómono, estableciendo: "*La respuesta es muy sencilla: la culpa será "a priori" del coche con sistema de asistencia a la conducción. Así de sencilla y concreta es la postura oficial de Volvo ante una de las cuestiones más difíciles de responder con total objetividad*. (…) *Sabemos que eso no tiene por qué ser así, pero para evitar suspicacias innecesarias, desde Volvo se plantea la máxima de que la responsabilidad es del coche.*"[399]

No obstante, respecto del accidente de marzo de 2018, la compañía alegó que la conducción autónoma del vehículo se hallaba bajo la supervisión del conductor de UBER, habida cuenta de que el sistema aun se encontraba en periodo de pruebas y, por ende, se entendía que la responsabilidad era de este (aunque también se abrió la posibilidad de que fuera UBER, como responsable del control y la gestión del servicio), que justamente tenía encomenda una tarea de control para actuar en caso de error y evitar un accidente. Finalmente, tal y como ya se había adelantado en su informe preliminar, la National Transport Safety Board Office of Highway Safety (NTSB)[400] llegó a la conclusión de que la causa del accidente fue un fallo en la programación del *software* del vehículo, habida

[398] Véase Comisión Europea, 2016.
[399] Volvo, 2016.
[400] Véase National Transport Safety Board Office of Highway, 2018.

cuenta de que este no estaba programado para detectar y actuar ante la aparición de peatones imprudentes que cruzaban por zonas no habilitadas para ello.

Por otro lado, en el caso del accidente del 23 de marzo ocurrido en California (EEUU), el conductor de un vehículo autónomo -frabricado por la compañía TESLA-, que al parecer llevaba conectado el sistema de piloto automático, perdió la vida como consecuencia de una colisión.

Ya en 2016, el Consejero Delegado de TESLA, Elon Musk, anunció que los vehículos creados por tal compañía irían equipados con un *hardware* que les permitiría ser totalmente autónomos, si bien advirtió de que la compañía no se haría legalmente responsable de los eventuales daños causados por estos, ya que según su punto de vista, ello debería ser asumido por la compañía de seguros contratada por cada propietario, salvo que estos se provocaran por un fallo endémico de diseño o fabricación, en cuyo caso la compañía sí asumiría su responsabilidad.[401]

No obstante, respecto del accidente de marzo de 2018, TESLA afirmó que el sistema de piloto automático empleado fue activado por el conductor, quien debía permanecer con atención al volante por si ocurría cualquier imprevisto, lo que hacía que el sistema no fuera completamente autónomo y, por ende, no resultara posible imputarse responsabilidad alguna a la compañía. No obstante, si bien al parecer, el conductor había estado jugando un juego en su teléfono durante el viaje, no estaba claro si también lo hacía momentos previos al accidente, según la investigación de la National Transportation Safety Board Office of Highway Safety.[402]

En relación con ello, numerosas iniciativas normativas específicas están surgiendo en el mundo en relación con los vehículos autónomos, si bien al ser un ámbito completamente ajeno al de esta tesis doctoral, no ha lugar a detenerse a analizar la misma en detalle.

En el ámbito civil, no obstante, resulta interesante hacer referencia a la ya mencionada Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones

[401] Muoio, 2016.
[402] Véase Chokshi, 2020.

destinadas a la Comisión sobre normas de Derecho civil sobre robótica, que considera que es crucial regular la cuestión de la responsabilidad jurídica por los daños que pueda ocasionar la actuación de los robots, habida cuenta de que, en materia de responsabilidad extracontractual, podría no ser suficiente el marco ofrecido por la Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados Miembros en materia de responsabilidad por los daños causados por productos defectuosos, que solo cubre los daños ocasionados por los defectos de fabricación de un robot a condición de que el perjudicado pueda demostrar el daño real, el defecto del producto y la relación de causa a efecto entre el defecto y el daño (responsabilidad objetiva o responsabilidad sin culpa); y que, pese al ámbito de aplicación de la Directiva 85/374/CEE, el marco jurídico vigente no bastaría para cubrir los daños causados por la nueva generación de robots, en la medida en que se les puede dotar de capacidades de adaptación y aprendizaje que entrañan cierto grado de imprevisibilidad en su comportamiento, ya que un robot podría aprender de forma autónoma de sus experiencias concretas e interactuar con su entorno de un modo imprevisible y propio únicamente de ese robot.[403]

Con posterioridad a ello, la Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de IA y robótica, que acogió con satisfacción la iniciativa de la Comisión Europea de crear un grupo de expertos sobre responsabilidad y nuevas tecnologías con el fin de proporcionar a la UE conocimientos especializados sobre la aplicabilidad de la Directiva relativa a la responsabilidad por productos defectuosos a los productos tradicionales, a las nuevas tecnologías y a los nuevos retos sociales, y señaló que los ingenieros de IA o las empresas que la emplean deben seguir asumiendo la responsabilidad de los impactos sociales, medioambientales y sanitarios que los sistemas de tal tecnología o la robótica puedan generar a las generaciones presentes y futuras.

En 2019, asimismo, se publicó por el Grupo de expertos en responsabilidad y nuevas tecnologías de la Comisión Europea un informe titulado: "*Liability for AI and other emerging digital technologies*"[404], que examinó los retos que surgen en torno a la

[403] Considerandos AH y AI.
[404] Comisión Europea, 2019.

regulación de la responsabilidad por los daños causados por los sistemas de IA y determinó que, si bien la normativa ya existente ofrece ciertas soluciones, es necesario revisar algunos aspectos, no entendiéndose imprescindible, no obstante, dotar a de una personalidad jurídica propia a los sistemas autónomos.

Y, finalmente, hay que hacer mención al reciente Informe de la Comisión al Parlamento Europeo, al Consejo y al Comité Económico y Social Europeo titulado: "Informe sobre las repercusiones en materia de seguridad y responsabilidad civil de la Inteligencia Artificial, el internet de las cosas y la robótica"[405] publicado el 19 de febrero del 2020, que, en materia de responsabilidad civil, examina principalmente la Directiva sobre seguridad general de los productos, así como las aportaciones de los grupos de expertos pertinentes y las consultas con las partes interesadas y constata que "*si bien, en principio, las normativas en vigor de la Unión y nacionales en materia de responsabilidad civil pueden hacer frente a las vicisitudes jurídicas derivadas de las tecnologías emergentes, la dimensión y el efecto combinado de las dificultades que plantea la IA podrían dificultar la indemnización de las víctimas en todos los casos en que esté justificada. Por lo tanto, el reparto de los costes cuando se produce un daño puede ser injusto o ineficiente con arreglo a las normas actuales*" y, con el fin de corregir tal situación, sugiere la realización de algunos ajustes en la mencionada Directiva y en los regímenes de responsabilidad civil nacionales, a través de iniciativas de la UE sobre la base de un enfoque específico basado en el riesgo, teniendo en cuenta que las distintas aplicaciones de IA presentan riesgos distintos.

En concreto, las problemáticas que se detectan en el ámbito de la responsabilidad civil por daños causados por sistemas de IA y sus posibles soluciones, son:

-la complejidad de los productos, los servicios y la cadena de valor.

Y es que, por un lado, por ejemplo, la línea divisoria entre los productos y servicios ya no es tan nítida como antes, siendo que ambos conceptos están cada vez más relacionados. Así, dado que hay programas informáticos de muchos tipos y formas, las respuestas relativas a la clasificación de estos como servicio o como producto no siempre son claras, por lo que, aunque la definición de producto de la Directiva sobre responsabilidad por los

[405] Comisión Europea, 2020.

daños causados por productos defectuosos es amplia, se entiende que debería precisarse su ámbito de aplicación para garantizar que haya una indemnización por los daños causados por productos defectuosos debido a sus programas informáticos u otras características digitales. Y, por otro lado, siendo que las aplicaciones de IA suelen estar integradas en entornos del "Internet de las cosas" complejos, en los que interactúan muchos dispositivos y servicios conectados, puede resultar difícil analizar y evaluar dónde se produce el perjuicio, quién es el responsable, y probar todas las condiciones que el Derecho nacional exige para la concesión de las indemnizaciones, por lo que procede analizar si podría ser una buena idea la reducción o inversión de la carga de la prueba exigida por las normas nacionales en materia de responsabilidad civil por los daños causados por el funcionamiento de las aplicaciones de IA, a través de una iniciativa de la UE;

-la autonomía y la opacidad.

Y es que para comprender un algoritmo y los datos utilizados por este hacen falta unos conocimientos técnicos que pueden resultar excesivamente gravosos para las víctimas y, de hecho, sin la cooperación de la parte aparentemente responsable, puede resultar imposible acceder a tal información. Además, no está claro cómo demostrar la culpa de un sistema de IA que ha actuado de forma autónoma, por lo que, de forma coordinada con los cambios correspondientes en el marco en materia de seguridad de la UE, la Comisión cree que debería revisarse el concepto de "puesta en circulación" empleado por la Directiva sobre responsabilidad por los daños causados por productos defectuosos, con el objetivo de tener en cuenta que los productos pueden cambiar y ser modificados, lo que sin duda podría ayudar a aclarar quién resulta el responsable civil de los cambios introducidos en los mismos.

En relación con el funcionamiento de las aplicaciones de IA con un perfil de riesgo específico, la Comisión está analizando si puede ser necesario, y en qué medida, por un lado, establecer una responsabilidad civil objetiva, a fin de indemnizar eficazmente a las posibles víctimas, y por otro lado, la posibilidad de vincular tal cuestión con la obligación de suscripción de un seguro, con el objetivo de garantizar el pago de la indemnización. Y en relación con el funcionamiento de las demás aplicaciones de IA, que son la gran

mayoría, la Comisión está reflexionando sobre si procede modificar y revisar la regulación relativa a la carga de la prueba de la causalidad y la culpa.

Así, si bien todavía no existe una normativa específica en el ámbito de la responsabilidad civil por los daños causados por sistemas de IA, lo cierto es que es un asunto que preocupa cada vez más y que está sobre la mesa, resultando evidente que la UE, al menos, está focalizando sus esfuerzos en establecer una regulación que esté a la altura de las circunstancias (lo cual no es nada fácil), consultando con expertos y tratando de aprovechar, en la medida de lo posible, la legislación existente, para garantizar una eficiente respuesta a las posibles víctimas de estos. Si bien existen debates sobre la posibilidad de que los sistemas de IA y los robots cuenten con una especie de personalidad jurídica denominada "personalidad electrónica", que les haría sujetos de derechos y obligaciones de forma autónoma, lo cierto es que, tal y como se deduce del documento analizado, la Unión Europea no se inclina por tal opción.

En el ámbito penal, por su parte, existen diversos estudios, sobre todo en relación con la eventual responsabilidad penal del vehículo autónomo, si bien es especialmente interesante hacer referencia al análisis realizado por John Kingston, investigador de la Universidad de Brighton (Reino Unido), sobre el trabajo efectuado por el también investigador de la Ono Academic College (Israel) Gabriel Hallevy, en que se ponen de manifiesto tres escenarios distintos relativos a la posible responsabilidad penal que podría derivarse en caso de comisión de un delito mediante el uso de un sistema de IA.[406]

En primer lugar, se hace mención al escenario denominado "perpetrador por medio de terceros", que hace referencia a aquellos casos en que un sujeto con conciencia influye e interviene en una persona que carece de capacidades intelectivas y volitivas para la comisión de un delito, valiéndose de él como mero "intermediario". En el ámbito de la IA, en concreto, podría entenderse que el sistema de tal tecnología es ese mero "intermediario" y que, los realmente responsables penalmente, por ende, son aquellos sujetos (personas físicas o incluso jurídicas) que lo programan o emplean para delinquir.

[406] Véase Kingston, 2018.

En segundo lugar, se hace mención al escenario denominado "consecuencia natural probable", que hace referencia a aquellos casos en que el uso inapropiado de las características y posibilidades normales de un sistema de IA provoca la comisión de un delito, resultando la pregunta clave en tal caso si el programador del mismo era conocedor de que tal resultado era una consecuencia probable de su uso o no. Como ejemplo se hace referencia a un robot inteligente que, en una fábrica en Japón, identificó erróneamente a un trabajador como una amenaza y trató de eliminarla dándole un empujón con su pesado brazo hidráulico, que aplastó al empleado y le causó la muerte en el instante, lo que sin duda es muestra evidente de producción de un resultado no deseado y contrario a la originaria intención de la creación del sistema de IA, pero ¿era una consecuencia naturalmente probable?.

Y en tercer lugar, se hace mención al escenario denominado "responsabilidad directa", que requiere acción e intención y abre la puerta a la responsabilidad penal del propio sistema de IA.

En mi opinión, el grado de autonomía del sistema de IA es el que deberá determinar el grado de responsabilidad, tanto civil como, en su caso, penal, del mismo. Y es que, está claro que si este ha sido diseñado para actuar sin supervisión humana, por ejemplo, cualquier fallo deberá atribuirse a su diseñador o, en su caso, fabricante, o incluso a los dos, en función de lo que un perito independiente concluya sobre el origen de tal error (puesto que no es lo mismo un incorrecto diseño del *software* que un defecto de fabricación, debiendo determinar quién habría podido haber previsto y evitado el resultado dañoso). Una buena opción para aumentar el rango de posibles responsables, sería la de considerar la posibilidad de permitir reclamar (solo civilmente) frente al comercializador del sistema que, no obstante, contaría con la facultad de repetir contra el diseñador, el fabricante o ambos. No obstante, si el sistema ha sido diseñado para actuar con supervisión humana en última instancia, salvo que exista un fallo de diseño o fabricación que perturbe ese control (que evidentemente deberá ser asumido por el diseñador o fabricante, tal y como ocurre hoy en día en el caso de los vehículos, por ejemplo), lo cierto es que bajo mi punto de vista, el que deberá ser considerado último responsable es el usuario, que en todo caso deberá tener contratado un seguro de responsabilidad civil obligatorio para asegurar que las eventuales víctimas ven reparados de forma efectiva los daños causados.

Y es que a mi parecer, otorgar personalidad jurídica propia a los sistemas de IA no es una solución ni deseable ni viable, ya que ello implicaría, por un lado, la aceptación de la existencia de sistemas completamente autónomos, cuando lo que para mí hay que garantizar y defender es la creación de sistemas que, en última instancia, puedan ser controlados por el ser humano, para evitar justamente que esta tecnología "se vaya de las manos"; y, por otro lado, abriría la puerta a la relajación de los estándares de seguridad de diseñadores y fabricantes de sistemas, que al saber que la responsabilidad recaería sobre las "máquinas", podrían verse despojados de presión y, por ende, decidir arriesgarse a lanzar sistemas sin cumplir con verificaciones o procesos de comprobación tan completos como en otros casos, evidentemente con el fin de reducir costes. Además, surge la pregunta de quién estaría detrás (económicamente hablando) de los sistemas de IA con personalidad jurídica autónoma para garantizar una respuesta eficaz por los eventuales daños causados. En relación con esto último, la ya mencionada Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, prevé la creación de un fondo de responsabilidad, a través de un sistema de seguros obligatorios, o bien de un fondo de compensación, incluso combinables entre sí, pero en mi opinión ello no soluciona el problema, puesto que, por un lado, tampoco está claro de dónde deberían salir los recursos para dotar de patrimonio a tales fondos ni quién debería contratar el seguro obligatorio y, por otro lado, ello no solventa los riesgos a los que anteriormente se ha hecho referencia.

3.- IA: HERRAMIENTA DE INVESTIGACIÓN CRIMINAL

3.1. LA IA Y EL PROCESO PENAL DE INSTRUCCIÓN ESPAÑOL

En España, son varios los instrumentos legales internacionales, europeos y nacionales que aplican y que deben ser respetados en el ámbito de la protección de derechos y, en concreto en su relación con el uso de las nuevas tecnologías. Así, y sin perjuicio de los convenios bilaterales y multilaterales suscritos con otros Estados, deben tenerse en cuenta: la Declaración Universal de los Derechos Humanos de la ONU, la Convención Europea de Derechos Humanos del Consejo de Europa, la Carta de Derechos Fundamentales de la UE, y la Constitución Española de 1978, vigente hoy.

Más en concreto, es importante remarcar, no obstante, que los únicos instrumentos generales que resultan jurídicamente vinculantes en nuestro país son: cuando se aplica el Derecho de la UE, la mencionada Carta de Derechos Fundamentales de la UE; y cuando se aplica Derecho nacional, la Constitución Española, sin perjuicio, por supuesto, de las leyes aplicables en cada caso concreto.

Es lógico, no obstante, que en tales normas no se haga una especial y amplia mención a las nuevas tecnologías (mucho menos a la IA) y su relación con los derechos fundamentales, habida cuenta de que en el momento en que ambas se aprobaron estas no tenían en nuestras sociedades la enorme incidencia que ostentan en la actualidad.

A pesar de ello, resulta necesario poner de relieve el novedoso y premonitorio artículo 18.4 de la Constitución Española de 1978, vigente hoy, que dispone: "*La ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos*.", siendo tal precepto el único de nuestra Carta Magna que hace referencia de forma específca al uso de las nuevas tecnologías y a su posible incidencia con los derechos fundamntales.

Por su parte, la Carta de Derechos Fundamentales de la UE, que no hace mención alguna al uso de la informática o las nuevas tecnologías de forma concreta, en su artícuo 8 sí

garantiza el derecho a la protección de los datos de carácter personal, cuestión fundamental en la era del *big data*. En relación con ello, si bien nuestra Constitución no contempla de forma específica un derecho a la protección de datos personales en la misma forma que la mencionada Carta de la UE, el Tribunal Constitucional ha sido el encargado de ir prefilando el contenido de tal derecho como una entidad autónoma, distinta del derecho a la intimidad.[407]

Tales normas básicas y supremas, no obstante, han sido desarrolladas por numerosos instrumentos legales dictados en el ámbito de la UE y de España, tal y como se ha ido haciendo referencia a lo largo de los puntos anteriores y tal y como se irá especificando en el ámbito concreto de cada herramienta de IA que posteriormente se analizará.

De forma específica, además, la específica relación de la IA y la justicia ha sido contemplada en algunas iniciativas y elementos normativos que, desde luego, deben tenerse en cuenta.

Por un lado, en el ámbito del Consejo de Europa, el 3 de diciembre de 2018, la Comisión Europea para la Eficacia de la Justicia (CEPEJ) adoptó la primera Carta Ética Europea sobre el uso de la IA en los sistemas judiciales[408], que incluía un estudio científico, un glosario y un análisis de distintas aplicaciones de IA con recomendaciones sobre su uso, estableciendo cinco principios que deberían guiar el desarrollo de las herramientas de IA en el ámbito de las Administraciones de Justicia europeas, a saber:

-principio de respeto de los derechos fundamentales, con garantía de que el diseño y la implementación de las herramientas y los servicios de IA sean compatibles con estos;
-principio de no discriminación, debiendo específicamente evitar el desarrollo o la intensificación de cualquier trato desigual entre individuos o grupos de individuos;
-principio de calidad y seguridad, en relación con el procesamiento de decisiones y datos judiciales, con garantía de uso de fuentes certificadas y modelos elaborados de forma multidisciplinaria, en un entorno tecnológico seguro;

[407] Véase, entre otras, la STC 292/2000, de 30 de noviembre.
[408] Consejo de Europa, 2018.

-principio de transparencia, imparcialidad y equidad, fomentando que los métodos de procesamiento de datos sean accesibles y comprensibles, con elaboración de auditorías externas; y
-principio de control del usuario, con exclusión de un enfoque imperativo y garantía de que los usuarios sean actores informados y que controlen las elecciones realizadas.

La idea de esta iniciativa fue que se desplegaran esfuerzos para garantizar que la Carta se convirtiera en "un instrumento vivo" dentro de los poderes judiciales europeos, de acuerdo con las orientaciones proporcionadas en el documento CEPEJ-GT-QUAL (2019) y, para ello, la Secretaría del CEPEJ se comprometió a organizar, a solicitud de los Estados miembros, actividades específicas para facilitar su implementación[409], lo que ya fue llevado a cabo con éxito en Reino Unido, habiéndose acordado que el CEPEJ fuera auditado por la asociación de profesionales "*The Law Society of England and Wales*".

Además, el Comité Europeo de Cooperación Jurídica (CDCJ) está trabajando en los mecanismos de resolución de disputas *on line*.

Por otro laso, en el ámbito de la UE, el 20 de febrero del 2020, en el seno del Comité sobre las Libertades Civiles del Parlamento Europeo, se celebró una reunión sobre el uso de la IA por parte de la policía y de las autoridades judiciales, en la que participaron representantes del Consejo de Europa, del Instituto de Investigación Interregional de Crimen y Justicia de las Naciones Unidas (UNICRI), de la Agencia de los Derechos Fundamentales de la Unión Europea, así como el Supervisor Europeo de Protección de Datos (SEPD), y grupos de expertos y representantes de la academia y de la sociedad civil, habiéndose centrado en los beneficios y riesgos de tal tecnología en el marco del Derecho Penal y las implicaciones éticas y de derechos fundamentales que ello podía conllevar.

Asimismo, el 8 de octubre de 2020 el Consejo de la Unión Europea publicó, en el marco del fin de la Presidencia alemana, un documento titulado "*Council Conclusions: Access to Justice-Seizing the Opportunities of Digitalisation*" [410], que alentaba a los Estados miembros a hacer un mayor uso de las herramientas digitales en los procedimientos

[409] Véase Consejo de Europa, 2019.
[410] Consejo de la Unión Europea, 2020.

judiciales, se solicitaba a la Comisión Europea que desarrollara una estrategia global de la UE sobre la digitalización de la justicia para finales de 2020, y se alegaba que el uso de las nuevas tecnologías, en especial la IA, no debía hacer tambalear los principios fundamentales de los sistemas judiciales, siendo necesario, no obstante, promover las habilidades digitales de jueces, fiscales, personal judicial y otros profesionales para que emplearan las herramientas tecnológicas de forma efectiva y con el debido respeto a los derechos y libertades de los ciudadanos.

Y el 21 de abril de 2021 se publicó, por fin, la esperada Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión. Tal texto, si bien todavía tiene que pasar el filtro del Consejo y del Parlamento Europeo y, por ende, no es definitivo, sin duda va a suponer una auténtica revolución en materia de IA en el ámbito de la UE y va a tener incidencia en lo relacionado con su uso en el ámbito judicial, tal y como se irá viendo en las próximas páginas.

Finalmente, en el ámbito español, por el momento la única normativa que contempla el uso de la IA en la investigación de delitos por parte de las autoridades (entre ellas, fiscales y judiciales) de forma específica es la LO 7/21, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales. No obstante, el 15 de abril de 2020 se anunció que el Ministro de Justicia iniciaba un proceso para la reforma del Real Decreto de 14 de septiembre de 1882 por el que se aprueba la Ley de Enjuiciamiento Criminal, (en adelante, LECrim), con el objetivo, entre otros, de introducir nuevos medios de investigación tecnológica y nuevas garantías en materia de protección de datos y derechos digitales.[411]

Resulta interesante al respecto, no obstante, lo establecido en la Exposición de Motivos de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas, que dispone (con subrayado propio): "*Si bien la Ley 30/1992*,

[411] Europa Press, 2020.

de 26 de noviembre, ya fue consciente del impacto de las nuevas tecnologías en las relaciones administrativas, fue la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos, la que les dio carta de naturaleza legal, al establecer el derecho de los ciudadanos a relacionarse electrónicamente con las Administraciones Públicas, así como la obligación de éstas de dotarse de los medios y sistemas necesarios para que ese derecho pudiera ejercerse. Sin embargo, en el entorno actual, la tramitación electrónica no puede ser todavía una forma especial de gestión de los procedimientos sino que debe constituir la actuación habitual de las Administraciones. Porque una Administración sin papel basada en un funcionamiento íntegramente electrónico no sólo sirve mejor a los principios de eficacia y eficiencia, al ahorrar costes a ciudadanos y empresas, sino que también refuerza las garantías de los interesados. En efecto, la constancia de documentos y actuaciones en un archivo electrónico facilita el cumplimiento de las obligaciones de transparencia, pues permite ofrecer información puntual, ágil y actualizada a los interesados.", que abre la puerta, sin duda, a la existencia de una Administración de justicia electrónica en la que, claramente, puede (y debe) tener cabida la IA.

Y es que en el ámbito de la investigación de delitos, la IA tiene un enorme potencial, habida cuenta de su capacidad para predecir tendencias de criminalidad y ayudar así a la optimización de recursos, facilitar la identificación de personas de interés, favorecer el hallazgo de vehículos u otros objetos robados, posibilitar la detección de comportamientos sospechosos, analizar datos de forma masiva y detectar fraudes económicos, corrupción o incluso actividades de financiación de grupos terroristas; facilitar la detección del uso y la distribución de material de pornografía infantil; favorcer el rastreo de redes de tráfico de personas, etc. No obstante, tal y como se ha ido poniendo de manifiesto a lo largo de la presente tesis doctoral, su uso no está exento de polémicos desafíos y retos, y bien es sabido que garantizar el respeto de los derechos humanos es una exigencia particularmente importante en el ámbito de la investigación criminal.

El procedimiento penal español, de naturaleza acusatoria, tal y como se recoge en la LECrim, cuenta con tres fases (con matices) muy diferenciadas: la fase de instrucción, la fase intermedia y la fase de plenario (con posterior eventual ejecución), todas ellas dirigidas

por el juez o magistrado (de instrucción y de lo penal, respectivamente), con asistencia de la Policía Judicial y del Ministerio Fiscal.

No obstante, existen diversos tipos o clases de procedimientos que otorgan especialidades a las mencionadas fases, si bien el periodo de instrucción suele tener eminentemente el mismo esquema en todos ellos.

Así, en virtud de lo dispuesto en los artículos 306 y siguientes de la LECrim, cuando al juez de instrucción le llega una *notitia criminis* (entendiéndose por tal una información relativa a la posible comisión de un delito), bien por vía de atestado policial, de denuncia, de querella, de informe remitido por un organismo público etc, tiene la obligación de, como mínimo, analizar su contenido para decidir si considera procedente y conforme a Derecho iniciar una investigación judicial o no.

En caso de tomar la decisión de dar comienzo a dicha investigación, se inicia la fase de instrucción del procedimiento penal, que en la mayoría de casos va precedida por una investigación policial (o de la fiscalía, en virtud de lo dispuesto en el artículo 773.2 de la LECrim) previa más o menos profunda, según los supuestos, que ayuda al juez a centrar el contenido y el objeto de las pesquisas.

El punto de partida para el juez de instrucción, con carácter general, sin duda, es el artículo 299 LECrim que dispone: "*Constituyen el sumario las actuaciones encaminadas a preparar el juicio y practicadas para averiguar y hacer constar la perpetración de los delitos con todas las circunstancias que puedan influir en su calificación y la culpabilidad de los delincuentes, asegurando sus personas y las responsabilidades pecuniarias de los mismos*."

En virtud de ello, el mencionado juez instructor, en ocasiones de forma conjunta con la policía y/o con el Ministerio Fiscal, decide qué diligencias, mínimas e imprescindibles (para evitar dilaciones indebidas, circunstancia poco deseable por estar prevista como atenuante en el artículo 22 de la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal), resultan pertinentes, útiles y adecuadas para proceder a la investigación de los presuntos hechos delictivos que han llegado a su conocimiento.

Tales diligencias, cuyas posibilidades son muy amplias y variadas dependiendo del/los delito/s que se pretenda/n investigar, suelen consistir en la toma de declaración de la presunta víctima, a la que se hace el oportuno ofrecimiento de acciones previsto en los artículos 109 y 110 LECrim, de los testigos y de la persona investigada, previa recabación de sus antecedentes penales, así como en el análisis de la documentación relacionada con el caso que pueda resultar de utilidad (y en algunas ocasiones, también, de imágenes o grabaciones de voz), la solicitud de información a otras Administraciones públicas o a empresas privadas, la adopción de medidas de seguimiento y vigilancia, de intervención de las comunicaciones, etc.

No obstante, por desgracia, ante la más que evidente insuficiencia de medios que acecha al ámbito de la justicia, tales investigaciones en muchas ocasiones devienen largas e ineficientes, y ello a pesar del plazo máximo de instrucción de 12 meses (con prórrogas sucesivas por periodos iguales o inferiores a 6 meses, en su caso) previsto en el artículo 324 LECrim, del que suelen acordarse prórrogas. Y es que muchas causas penales pasan por manos de distintos jueces de instrucción, fiscales y funcionarios desbordados que apenas pueden dedicarles tiempo, se topan con dificultades para practicar las diligencias acordadas (imposibilidad de hallar a la persona investigada, falta de candidatos idóneos para la realización de ruedas de reconocimiento, volúmenes de documentación absolutamente ingentes y abrumadores que requieren para su análisis una capacidad y un tiempo de los que el juez de instrucción carece, problemas para hallar peritos expertos en ciertas materias dispuestos a aceptar la realización de informes periciales que les resultan poco rentables, comisiones rogatorias cuyos resultados tardan años en llegar, etc), e incluso con obstáculos que impiden seguir adelante con la instrucción del caso. No obstante, hoy en día ya disponemos de herramientas tecnológicas, en concreto de IA, que son las que constituyen el objeto de la presente tesis doctoral, que facilitarían mucho el trabajo de los profesionales de la justicia y ayudarían, sin duda, a ofrecer un servicio de mayor calidad a los ciudadanos, lo que bajo mi punto de vista debe ser la máxima aspiración de todo servidor público.

En relación con ello, piénsese por ejemplo, en una persona mayor que ha sido víctima de una estafa en su domicilio por personas que se hacían pasar por técnicos de la luz o del gas y cobraban por trabajos innecesarios y mal realizados. En algunas ocasiones, tales fraudes,

una vez se detecta que la víctima "cae en la trampa", devienen continuados y llegan a crear auténticos destrozos en la economía de los afectados. No obstante, cuando la *notitia criminis* llega al juez de instrucción, en aquellos casos en que, tras un arduo trabajo de investigación de la policía se ha conseguido identificar a los posibles autores, llega el momento de someterlos a una rueda de reconocimiento para resultar, en su caso, reconocidos por la/s víctima/s. ¿El problema? Por un lado, en la mayoría de casos (mal que nos pese, consecuencia de la sobrecarga de la agenda de los juzgados, de la imposibilidad de hallar figurantes similares a los autores de los hechos o de la incomparecencia -a veces reiterada- de los citados), tal diligencia de investigación se practica tanto tiempo después de la comisión de los hechos que a las víctimas les resulta prácticamente imposible recordar el rostro e identificar a los presuntos autores; y, por otro lado, a pesar de que pase poco tiempo entre una acción y otra, en caso de que la/s persona/s afectada/s sea/n mayor/es o incapaz/ces, haya/n sufrido una gran situación de estrés en el momento de los hechos, o sea/n acechada/s por un sentimiento de miedo o angustia ante la práctica de tal diligencia de investigación, las posibilidades de que logre/n identificar con seguridad al/los presunto/s estafador/es son mínimas, lo que conlleva que en aquellos casos en que no existan otros indicios lo suficientemente fuertes como para hacer tambalear la presunción de inocencia y seguir adelante con la investigación, estos hechos (en ocasiones graves) queden impunes.

Ello, no obstante, con un sistema de reconocimiento facial potente y fiable (IA), podría, como veremos más adelante, solventarse.

Asimismo, piénsese en un caso de presuntos delitos de administración desleal, estafa y blanqueo de capitales que llega al juzgado por querella presentada por una empresa afectada que aporta, como documental, las cuentas anuales de los últimos cuatro años de varias sociedades relacionadas entre sí, movimientos de decenas de cuentas bancarias, un informe pericial de más de quinientos folios, una solicitud de oficiar a seis entidades financieras diferentes para que aporten información de las cuentas bancarias de quince personas distintas presuntamente implicadas, y de efectuar tres comisiones rogatorias (una a Suiza, otra a Islas Bahamas y otra a Sudáfrica) para obtener datos de las posibles transacciones internacionales efectuadas. ¿Qué mente humana, sin preparación técnica sobre el asunto y sin tiempo apenas para dedicar al estudio, puede analizar toda la documentación presentada y obtenida, interpretar y entender de la forma más objetiva

posible el informe pericial aportado por la acusación particular, el más que probable informe pericial presentado por la defensa para contradecirlo, y el eventual informe pericial solicitado de oficio a un perito judicial, y dar una respuesta eficiente y de calidad al ciudadano que solicita asistencia? Ello, sin duda, es una ardua tarea que en muchas ocasiones desborda al fiscal y al juez de instrucción encargados del asunto (que, como suele decirse "son humanos"), lo que se traduce en investigaciones que se eternizan de forma "injustificada", en la posterior apreciación de la circunstancia atenuante de dilaciones indebidas, y en algunos casos, incluso, en el archivo de la causa por prescripción.

Ello, no obstante, con un sistema de análisis de documentos y cruce de datos potente y fiable (IA), podría, como veremos más adelante, mejorar considerablemente.

Piénsese también en el caso de un grupo de ciudadanos españoles que tiene como objetivo la comisión de delitos contra la propiedad (en concreto, robos con fuerza en domicilios). Imagínese que la policía hace un gran trabajo, recibe informaciones, realiza vigilancias, investiga y consigue detener *in fraganti* a los autores de los hechos, momento en que estos pasan a disposición judicial y son dejados en libertad provisional con cargos habida cuenta de que tienen arraigo en nuestro país y carecen de antecedentes penales. ¿Qué es posible que ocurra -y de hecho, ocurre en numerosas ocasiones-, habida cuenta de que la pena prevista para el delito de robo con fuerza en casa habitada, en virtud de lo dispuesto en el artículo 241 del CP, es de dos a cinco años de prisión? Pues que los autores de los hechos den direcciones y teléfonos que no se corresponden con la realidad o den datos ciertos pero de forma inmediatamente posterior dejen de atender a los requerimientos del juzgado o incluso modifiquen sus domicilios, trasladándose hasta de ciudad sin dejar rastro, y aun cuando queden obligados a comparecer *apud acta* cada X días en el juzgado más próximo a su domicilio, devengan ilocalizables para la Administración de Justicia. En tales casos, y tras realizarse por el Letrado de la Administración de Justicia, y en ocasiones por la policía, las oportunas -muy básicas- averiguaciones de domicilio y de algun otro dato de localización más al que se pueda acceder, suele acordarse sobre ellos una orden de busca y captura que, salvo en contadas ocasiones (por la comisión de nuevos hechos delictivos que comporten actuación policial, por la identificación casual en un control policial, por la entrada en un hotel con aportación de sus datos personales, etc), acaba conllevando el archivo de la causa por prescripción, por imposibilidad de hallar a los huidos.

Ello, con el uso de gafas inteligentes (IA) dotadas de sistemas de reconocimiento facial por parte de los Cuerpos y Fuerzas de Seguridad del Estado, por ejemplo, podría mejorarse exponencialmente.

Otro ejemplo de instrucción frustrada y mejorable sería, por ejemplo, el de en un caso que actualmente (cada día más) se halla en la mayor parte de juzgados de instrucción de España y del resto del mundo: el de las estafas por Internet. Y es que, en muchas ocasiones, tales fraudes se cometen de forma transnacional, a saber, con creación de una dirección IP espejo desde Israel, envío de un correo electrónico con virus desde no se sabe dónde, recepción de este y sustracción de dinero de una cuenta basada en España, y envío del beneficio a una cuenta bancaria registrada en China a nombre de una sociedad con base en las Islas Seychelles. ¿Qué ocurre en casos así? Pues, por desgracia, y a pesar de que los Cuerpos y Fuerzas de Seguridad españoles (tanto nacionales como autonómicas) cuentan con grandes profesionales expertos en la materia que se dedican sin descanso a seguir el rastro de tales operaciones para conseguir identificar a los posibles autores, en muchas ocasiones tanto estos como el juez de instrucción se topan con una clara limitación en las herramientas de que disponen para llevar a cabo la investigación. Y es que, en la mayoría de casos, se pierde la pista de los posibles autores en momentos muy iniciales de la investigación y, en otros, se acaba dependiendo de la respuesta de una comisión rogatoria remitida a países que, solo en contadas ocasiones, proporcionan información útil y completa (y en un periodo de tiempo tan corto que evite que el dinero depositado en las cuentas basadas en estos desaparezca). ¿Resultado? Muchas estafas cometidas por los medios expuestos acaban siendo archivadas en nuestros tribunales por imposibilidad de identificar y hallar a las personas (físicas o jurídicas) responsables de los hechos.

Ello, no obstante, con un sistema de trazabilidad y análisis y cruce de datos (IA) potente y, desde luego, una mayor cooperación internacional, podría llegar a solventarse.

En nuestra LECrim, a raíz de la reforma operada por la LO 13/2015, de 5 de octubre, de modificación de dicho cuerpo legal para el fortalecimiento de las garantías procesales y la regulación de las medidas de investigación tecnológica, se introdujo en el Título VIII, titulado “De las medidas de investigación limitativas de los derechos reconocidos en el artículo 18 de la Constitución”, el Capítulo IV, bajo el título “Disposiciones comunes a la

interceptación de las comunicaciones telefónicas y telemáticas, la captación y grabación de comunicaciones orales mediante la utilización de dispositivos electrónicos, la utilización de dispositivos técnicos de seguimiento, localización y captación de la imagen, el registro de dispositivos de almacenamiento masivo de información y los registros remotos sobre equipos informáticos", que a través de lo dispuesto en los artículos 588bis a) a 588 bis k), establece las bases comunes a tener en cuenta para la utilización de ciertos medios de investigación tecnológicos que en los posteriores capítulos se regulan de forma más específica y profunda.

Bajo mi punto de vista, tal regulación sería, sin duda, un buen punto de partida (con matices) aplicable a los medios de investigación criminal que emplean IA, habida cuenta de que, en todo caso, la posibilidad de vulnerar los derechos fundamentales en juego y la necesidad de garantizarlos es una necesidad común en el uso de tal tecnología y del resto de las ya reguladas.

En relación con ello, la base principal se sienta en el artículo 588 bis a de la LECrim, que en su apartado 1 dispone: "*Durante la instrucción de las causas se podrá acordar alguna de las medidas de investigación reguladas en el presente capítulo siempre que medie autorización judicial dictada con plena sujeción a los principios de especialidad, idoneidad, excepcionalidad, necesidad y proporcionalidad de la medida*."

Así, de entrada, ya se exigen dos requisitos mínimos para la utilización de cualquier medio de investigación tecnológico en la fase de instrucción, a saber: autorización judicial y sujeción de esa autorización a los principios de especialidad, idoneidad, excepcionalidad, necesidad y proporcionalidad, lo que se erige como *conditio sine qua non* para dotar de validez a la medida. A continuación, se expone a qué hacen referencia exactamente cada uno de tales principios, y en tal sentido se dispone que:

-el *principio de especialidad* exige que una medida esté relacionada con la investigación de un delito concreto, no pudiendo autorizarse medidas de investigación tecnológica que tengan por objeto prevenir o descubrir delitos o despejar sospechas sin base objetiva;

-el *principio de idoneidad* sirve para definir el ámbito objetivo y subjetivo y la duración de la medida, en virtud de su utilidad;

-en aplicación de los *principios de excepcionalidad y necesidad* solo podrá acordarse la medida:

a) cuando no estén a disposición de la investigación, en atención a sus características, otras medidas menos gravosas para los derechos fundamentales del investigado o encausado e igualmente útiles para el esclarecimiento del hecho, o

b) cuando el descubrimiento o la comprobación del hecho investigado, la determinación de su autor o autores, la averiguación de su paradero, o la localización de los efectos del delito se vea gravemente dificultada sin el recurso a esta medida.

-el *principio de proporcionalidad* implica que, tomadas en consideración todas las circunstancias del caso, el sacrificio de los derechos e intereses afectados no sea superior al beneficio que de la adopción de la medida resulte para el interés público y de terceros. En Para la ponderación de los intereses en conflicto, la valoración del interés público se basará en la gravedad del hecho, su trascendencia social o el ámbito tecnológico de producción, la intensidad de los indicios existentes y la relevancia del resultado perseguido con la restricción del derecho.

Además, el apartado 1 del artículo 588 bis c LECrim añade otro requisito: la audiencia del Ministerio Fiscal antes de dictar la resolución judicial que acuerde (de oficio o a instancia de la Policía Judicial) o deniege la medida correspondiente, que deberá contener, al menos, según lo dispuesto en el apartado 3, los siguientes extremos (con subrayado propio):

"*a) El hecho punible objeto de investigación y su calificación jurídica, con expresión de los indicios racionales en los que funde la medida.*
b) La identidad de los investigados y de cualquier otro afectado por la medida, de ser conocido.

c) La extensión de la medida de injerencia, especificando su alcance así como la motivación relativa al cumplimiento de los principios rectores establecidos en el artículo 588 bis a.
d) La unidad investigadora de Policía Judicial que se hará cargo de la intervención.
e) La duración de la medida.
f) La forma y la periodicidad con la que el solicitante informará al juez sobre los resultados de la medida.
g) La finalidad perseguida con la medida.
h) El sujeto obligado que llevará a cabo la medida, en caso de conocerse, con expresa mención del deber de colaboración y de guardar secreto, cuando proceda, bajo apercibimiento de incurrir en un delito de desobediencia."

Se establece, asimismo, en el apartado 1 del artículo 588 bis e de la LECrim, una limitación en cuanto a la duración de las medidas tecnológicas de investigación, habida cuenta de los derechos fundamentales que están en juego: no podrán exceder del tiempo imprescindible para el esclarecimiento de los hechos (del mismo modo que el periodo de detención policial -también limitativo de derechos fundamentales-, en virtud de lo dispuesto en el artículo 17.2 de la Constitución Española, no podrá durar más tiempo del estrictamente necesario para la realización de las averiguaciones tendentes al esclarecimiento de los hechos, a pesar de que, en tal supuesto se fija un plazo máximo de 72 horas, lo cual no ocurre en el caso de la duración de las medidas tecnológicas que, de hecho, pueden ser prorrogadas mediante auto motivado siempre que subsistan las causas que las justificaron, *ex* apartado 2 del artículo 588 bis e de la LECrim, con las únicas limitaciones temporales derivadas de lo dispuesto en el artículo 324 del mismo cuerpo legal).

Una particularidad relativa a la solicitud de las medidas de investigación tecnológica y posteriores actuaciones que se lleven a cabo en relación con las mismas es que estas deben sustanciarse en pieza separada y secreta, sin necesidad de acordar de forma expresa el secreto de las actuaciones, tal y como dispone el artículo 588 bis d LECrim.

Por su parte, en relación con la posible afectación a terceras personas como consecuencia de la adopción de las medidas de investigación tecnológica, el artículo 588bis h de la LECrim deja bien claro que estas solo podrán acordarse en los casos y con las condiciones

que se regulan en las disposiciones específicas de cada una de ellas, y ello con el fin de salvaguardar y respetar al máximo los derechos y libertades de los terceros, de forma ponderada, no obstante, con el fin de la investigación. Y, asimismo, con relación a utilización de la información obtenida en un procedimiento distinto y a los denominados "hallazgos casuales", el artículo 588 bis i de la LECrim se remite a lo dispuesto en el artículo 579 bis del mismo cuerpo legal, que autoriza el uso, como medio de investigación o prueba en otro proceso penal, del resultado obtenido mediante la adopción de las respectivas medidas de investigación tecnológicas, si bien requiere la deducción de testimonio de los particulares necesarios para acreditar la legitimidad de la injerencia y, para continuar con el uso de la medida para la investigación del delito casualmente descubierto, se requiere autorización del juez competente en los términos específicamente previstos en el apartado 3 del mencionado precepto.

Respecto del control de las medidas, se establece en el artículo 588 bis g de la LECrim la necesidad de supervisión judicial, para lo cual se dispone que la Policía Judicial informará al juez de instrucción del desarrollo y de los resultados de las mismas, en la forma y con la periodicidad que este determine y, en todo caso, cuando por cualquier causa se ponga fin a la misma,

Y, finalmente, en cuanto al cese de las medidas, por un lado, el artículo 588 bis j de la LECrim establece que el juez de instrucción lo acordará cuando desaparezcan las circunstancias que justificaron su adopción o cuando resulte evidente que, a través de las mismas, no se están obteniendo los resultados pretendidos y, en todo caso, cuando haya transcurrido el plazo para el que hubieran sido autorizadas; y, por otro lado, el artículo 588 bis k de la LECrim ordena que, una vez que se ponga término mediante resolución firme al procedimiento en que se emplearon medidas de investigación tecnológica, se lleve a cabo por la Policía Judicial el borrado y la eliminación de los registros originales que puedan constar en los sistemas electrónicos e informáticos utilizados en la ejecución de estas, con conservación de una copia bajo custodia del Letrado de la Administración de Justicia (que se destruirá cuando hayan transcurrido cinco años desde que la pena se haya ejecutado, cuando el delito o la pena hayan prescrito, se haya decretado el sobreseimiento libre, o haya recaído sentencia absolutoria firme respecto del investigado, siempre que no fuera precisa su conservación, a juicio del Tribunal).

De acuerdo con lo anterior, es evidente que en nuestra LECrim ya disponemos, tal y como se ha advertido, de una base regulatoria aplicable a las medidas de investigación tecnológica que, sin duda, se ajusta a los estándares de protección específica y garantista de los posibles derechos fundamentales en juego que exige la Constitución Española para el Derecho nacional. Y ello podría, sin duda, servir de regulación general para el uso de los sistemas de IA, que no obstante, se obvian por completo en la actual regulación.

No obstante, para poder determinar la legalidad de una herramienta de IA resulta fundamental que esta sea lo suficientemente transparente y explicable como para poder dilucidar, de forma real, los posibles derechos fundamentales afectados, ya que en caso contrario, tal tarea deviene enormemente dificultosa.

Ello obliga a traer a colación, una vez más, para evitar el exceso de judicialización de los asuntos y el consiguiente colapso de los tribunales, la necesidad de que el legislador otorgue unas pautas claras e inteligibles a los creadores de sistemas de IA, ya que así, de forma inicial, podrían dotar a estos de ciertas limitaciones que, sin duda, ayudarían a acotar el (buen) uso de los mismos con posterioridad.

Y, asimismo, resulta fundamental para este ámbito la -ya reclamada en numerosas ocasiones a lo largo de esta tesis doctoral- creación de una Agencia a nivel europeo y/o nacional que certifique la calidad de los sistemas de IA con carácter previo a su distribución y uso y, especialmente, garantice la transparencia y explicabilidad de los mismos, puesto que solo así podrá pasarse al siguiente escalón: el de realizar la ponderación de derechos, de forma fiel a la realidad. Y es que si bien es cierto que lo ideal y lo deseable, en su caso, sería que dicha Agencia, tal y como ya se ha avanzado en puntos anteriores, llevara a cabo un filtro previo de los sistemas asegurándose de que cumplen con los valores, principios y derechos en los que se basa el ordenamiento jurídico europeo, imprescindible para asegurar la mayor seguridad jurídica posible, lo cierto es que siempre podrían darse casos concretos de colisiones de derechos que habría que analizar de forma concreta con posterioridad por parte de los tribunales, a sabiendas de que la casuística es infinita.

Las buenas noticias, no obstante, están al caer. Y es que la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en

materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, establece una regulación específica para los sistemas de IA y, en concreto, para aquellos más empleados en el ámbito de la investigación criminal.

Así, con carácter general, y sin perjuicio de lo que se expondrá más adelante en relación con cada una de las concretas herramientas, en el Anexo III del mencionado texto legal se hace referencia a los sistemas de IA de alto riesgo, en consonancia con lo dispuesto en el artículo 6.2, y en especial se nombra a:

"*-los sistemas de identificación biométrica y categorización de personas físicas:*

a)sistemas de IA destinados a utilizarse en la identificación biométrica remota «en tiempo real» o «en diferido» de personas físicas.

(…)

-asuntos relacionados con la aplicación de la ley:

a) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para llevar a cabo evaluaciones de riesgos individuales de personas físicas con el objetivo de determinar el riesgo de quecometan infracciones penales o reincidan en su comisión, así como el riesgo para las potenciales víctimas de delitos;

b) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley como polígrafos y herramientas similares, o para detectar el estado emocional de una persona física;

c) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para detectar ultrafalsificaciones a las que hace referencia el artículo 52, apartado 3;

d) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para la evaluación de la fiabilidad de las pruebas durante la investigación o el enjuiciamiento de infracciones penales;

e) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para predecir la frecuencia o reiteración de una infracción penal real o potencial con base en la elaboración de perfiles de personas físicas, de conformidad con lo dispuesto en el artículo 3, apartado 4, de la Directiva (UE) 2016/680, o en la evaluación de rasgos y características de la personalidad o conductas delictivas pasadas de personas físicas o grupos;

f) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para la elaboración de perfiles de personas físicas, de conformidad con lo dispuesto en el artículo 3, apartado 4, de la Directiva (UE) 2016/680, durante la detección, la investigación o el enjuiciamiento de infracciones penales;

g) sistemas de IA destinados a utilizarse para llevar a cabo análisis sobre infracciones penales en relación con personas físicas que permitan a las autoridades encargadas de la aplicación de la ley examinar grandes conjuntos de datos complejos vinculados y no vinculados, disponibles en diferentes fuentes o formatos, para detectar modelos desconocidos o descubrir relaciones ocultas en los datos."

Tales sistemas, que sin duda pueden resultar de enorme utilidad en el ámbito de la investigación criminal, como se expondrá en los puntos siguientes, al ser considerados de "alto riesgo" quedan sujetos a una regulación muy estricta (que conlleva incluso la prohibición de su uso, en algunos casos, tal y como se desprende de lo dispuesto en el artículo 6) y, en concreto, en el Capítulo 2 del Título III se prevé la creación de un sistema de gestión de riesgos continuado en el tiempo, en los términos del artículo 9; se prevé la creación de un sistema de gobernanza de datos que asegure la calidad de los mismos, en los términos del artículo 10; se prevé la realización de registros constantes ("archivos de registro") mientras están en funcionamiento que garanticen un nivel de trazabilidad de este durante su ciclo de vida, permitan detectar y controlar la aparición de situaciones que puedan hacer que el sistema presente un riesgo o dar lugar a una modificación sustancial, y faciliten el seguimiento posterior a la comercialización al que se refiere el artículo 61, todo ello en los términos del artículo 12; se prevé una garantía extra de transparencia y comunicación de información a los usuarios, con obligación de proporcionar unas instrucciones de uso en formato digital o de otro tipo adecuado, en los términos del artículo 13; se prevé garantizar que tales sistemas puedan ser vigilados de manera efectiva por

personas físicas durante el periodo en que estén en uso, en los términos de lo dispuesto en el artículo 14; y se prevé garantizar un nivel muy elevado de precisión, solidez y ciberseguridad, con resistencia a los errores, fallos e incoherencias que puedan surgir en los propios sistemas o en los entornos donde operen, en particular a causa de su interacción con personas físicas u otros sistemas, en los términos previstos en el artículo 15.

Asimismo, y en el mismo sentido, se establecen una serie de obligaciones para los proveedores de tales sistemas calificados de "alto riesgo", en los términos previstos en el artículo 16. En relación con ello, entre otras, se les impone la obligación de establecer un sistema de gestión de la calidad que garantice el cumplimiento de lo dispuesto en el futuro Reglamento, en los términos previstos en el artículo 17; se les impone la obligación de asegurarse de que sus sistemas sean sometidos al procedimiento de evaluación de la conformidad oportuno, de acuerdo con lo previsto en el artículo 43, antes de su introducción en el mercado o puesta en servicio, y se dispone que, cuando dicha evaluación de conformidad demuestre que los sistemas de IA cumplen con los requisitos establecidos en el Capítulo 2 del Título III, los mencionados proveedores elaborarán una declaración UE de conformidad con arreglo a lo dispuesto en el artículo 48 y colocarán el marcado CE conforme a lo previsto en el artículo 49; se les impone la obligación de conservar los "archivos de registro" que generen automáticamente sus sistemas de IA en los términos previstos en el artículo 20; se les impone la obligación de adoptar, de forma inmediata, las medidas correctoras necesarias en caso de advertir o tener motivos para advertir que un sistema de IA que han introducido en el mercado o puesto en servicio no es conforme con lo dispuesto en el futuro Reglamento, en los términos dispuestos en el artículo 21; y se les impone la obligación de cooperar con las autoridades competentes en los términos establecidos en el artículo 23.

Por su parte, también se fijan obligaciones para los fabricantes, en los términos previstos en el artículo 24; para los importadores, en los términos dispuestos en el artículo 26; para los distribuidores, según lo establecido en el artículo 27; e incluso para los usuarios, que principalmente deberán usar estos sistemas con arreglo a las instrucciones de uso que acompañen, en los términos del artículo 29.

Y con el fin de garantizar que se llevan a cabo correctamente los procesos de evaluación, designación y notificación de los organismos de evaluación de la conformidad de los

sistemas de IA, así como su seguimiento, el Capítulo 4 del Título III establece unas normas específicas, delegando en los Estados Miembros el nombramiento o constitución de una autoridad notificante que será responsable de establecer y llevar a cabo todo lo necesario para ello.

Y, respecto de la evaluación de la conformidad con la legalidad vigente de los sistemas de IA de alto riesgo, el artículo 43 establece un procedimiento ordinario, si bien el artículo 47 prevé un procedimiento de urgencia, por razones excepcionales de seguridad pública o con el fin de proteger la vida y la salud de las personas, el medio ambiente y los activos fundamentales de las industrias e infraestructuras.

En relación con ello, el artículo 43 diferencia entre el procedimiento a seguir por los proveedores de los sistemas de IA de alto riesgo enumerados en el punto 1 del Anexo III, a saber, los sistemas de "*Identificación biométrica y categorización de personas físicas: a) sistemas de IA destinados a utilizarse en la identificación biométrica remota en tiempo real o en diferido de personas físicas*." y los enumerados en los puntos 2 a 8 del mencionado Anexo.

Respecto de los primeros, se contemplan dos opciones (con excepciones): que el proveedor lleve a cabo un procedimiento de evaluación de la conformidad fundamentado en un control interno, en los términos previstos en el Anexo VI; o que se someta a un procedimiento de evaluación de la conformidad con la participación de un organismo notificado en los términos previstos en el Anexo VII (en caso de que se prevea la puesta en servicio del sistema por parte de las autoridades encargadas de aplicar la ley o las instituciones, organismos o agencias de la UE, el orgaismo notificado será la autoridad de vigilancia del mercado mencionada en el artículo 63).

En el primer supuesto, el proveedor redactará una declaración UE de conformidad para cada sistema de IA que introduzca en el mercado o ponga en servicio y la mantendrá a disposición de las autoridades nacionales competentes durante un plazo de diez años en los términos previstos en el artículo 48. Y en el segundo supuesto, cuando el sistema de IA supere con éxito el procedimiento de evaluación de la conformidad, la idea es que se expida un certificado de calidad por parte de los organismos notificados conforme a lo dispuesto en el Anexo VII que será válido por un periodo no superior a cinco años (con posibilidad

de prórroga por periodos renovables no superiores a cinco años previa reevaluación de la conformidad). No obstante, en caso de que un organismo notificado observe que un sistema de IA ya no cumple con los requisitos legalmente establecidos, deberá suspender, limitar o retirar, con aplicación del principio de proporcionalidad, el certificado expedido, conforme a lo dispuesto en los artículos 44 y 46.

Respecto de los segundos, solo se contempla una opción: que los proveedores se atengan al procedimiento de evaluación de la conformidad fundamentado en un control interno en los términos previstos en el Anexo VI, sin contemplar la posibilidad de participación de un organismo notificado.

Con el fin de dar a conocer que un sistema de IA es conforme con la legislación vigente, tal y como prevé el artículo 49 contará con el marcado CE qe se colocará de modo visible y legible.

Y, antes de la introducción en el mercado o la puesta en servicio de tales sistemas de IA, el proveedor los registrará en la base de datos de la UE, creada por la Comisión, a la que se refiere el artículo 60.

Además de lo expuesto, la Propuesta de Reglamento incluye algunas medidas de gobernanza tales como la creación de un Comité Europeo de IA que, según lo dispuesto en el artículo 56, tendrá el fin de ofrecer asesoramiento y asistencia a la Comisión para contribuir a la cooperación efectiva entre esta y las autoridades nacionales de supervisión respecto de las materias reguladas; coordinar y contribuir a las orientaciones y los análisis de la Comisión y las autoridades competentes sobre problemas emergentes en el mercado interior respecto de las materias reguladas; y asistir a la Comisión y a las autoridades nacionales de supervisión para garantizar la aplicación coherente del Reglamento.

Y, asimismo, el artículo 59 prevé que cada Estado miembro establezca o designe autoridades nacionales competentes con el fin de garantizar la aplicación y la ejecución del futuro Reglamento, lo cual celebro, puesto que considero que es muy necesario que se vele por su cumplimiento, habida cuenta de la cantidad de derechos fundamentales que podrían verse vulnerados en caso contrario y, sobre todo, habida cuenta de que la credibilidad de la UE en materia de garantías de los sistemas de IA está en juego. Y no solo eso, sino que

además se prevén en el artículo 71 una serie de sanciones para casos de incumplimiento y se delega en los Estados miembros la creación de un régimen de sanciones para los supuestos de vulneración de lo establecido en el futuro Reglamento.

Finalmente, tal y como ya se ha expuesto, en el Título VIII se prevé un seguimiento posterior a la comercialización de los sistemas de IA, en los términos previstos en el artículo 61; y se impone a los proveedores de sistemas de alto riesgo la obligación de notificar a las autoridades de vigilancia del mercado cualquier incidente grave o fallo de funcionamiento que constituya un incumplimiento de las obligaciones de Derecho de la UE destinadas a proteger a los derechos fundamentales.

Y, a la vista de lo expuesto, que todavía es una mera propuesta y, por ende, no constituye legislación aplicable, queda patente la urgente necesidad de regular la introducción y puesta en servicio de sistemas de IA en el ámbito de la UE, especialmente de los calificados de alto riesgo, entre los que se encuentran, como ya he dicho con anterioridad, una buena parte de los que pueden resultar útiles en el ámbito de la investigación penal.

3.2. HERRAMIENTAS DE IA PARA INVESTIGAR DELITOS

A lo largo de los anteriores capítulos ha resultado necesario establecer y exponer las líneas más generales (conceptuales, históricas, regulatorias, etc) del significado y las circunstancias que rodean a la tecnología que da razón de ser a la presente tesis doctoral: la IA. Y es que, desde luego, resulta necesario para poder analizar de forma profunda la hipótesis que subyace en este trabajo, tener un conocimiento claro de cuáles son los desafíos a los que la tecnología analizada se enfrenta y, sobre todo, cuál es la verdadera base de la investigación que nos ocupa.

Sentado lo anterior, la presente Sección es, sin duda, la "estrella" de la presente tesis doctoral y desde luego el grueso del contenido de la misma. Y es que la pregunta "¿Cómo puede la IA ayudar o contribuir en la investigación de delitos en el procedimiento de instrucción español?" halla respuestas concretas únicamente en el análisis de todas aquellas herramientas que tal tecnología ofrece y, especialmente, de los retos y eventuales problemas jurídicos que estas plantean y sus posibles soluciones.

Para llevar a cabo un estudio completo y ordenado de las distintas herramientas de IA que pueden resultar de utilidad en el procedimiento de instrucción español, he creído conveniente establecer una división de las mismas en tres bloques, a saber:

-herramientas de predicción y evaluación de riesgos;
-herramientas de investigación de delitos propiamente dichas; y
-herramientas de tramitación.

Y es que se entiende necesario y adecuado llevar a cabo tal clasificación de las distintas herramientas de IA en los mencionados bloques puesto que, como se observará, guarda lógica agruparlas según su naturaleza y sus finalidades, ya que ello facilita y sistematiza enormemente su análisis, habida cuenta de que, sin perjuicio de las particulares características y posibles problemáticas que presentan cada una de ellas, existen elementos comunes de evaluación que pueden resultar muy útiles para varios de los sistemas e instrumentos que se van a examinar.

Empezamos.

3.2.1. HERRAMIENTAS DE PREDICCIÓN Y EVALUACIÓN DE RIESGOS

3.2.1.1 Concepto

Las herramientas de IA de predicción y evaluación de riesgos pueden ser definidas como aquellos sistemas que emplean tal tecnología con la finalidad de predecir eventos futuros y, por ende, valorar la existencia de posibles y potenciales peligros o riesgos venideros, a saber: dónde es más probable que vaya a producirse un crimen, si un detenido tiene riesgo de fuga o de reiteración delictiva, si un investigado por violencia de género es probable que vaya a atentar de nuevo contra la víctima, si es previsible que un preso reingrese en prisión tras la concesión de un permiso penitenciario, o si una empresa va a deshacerse de sus activos tras la notificación del inicio de su investigación, entre otros. Y es que, si bien

los algoritmos no pueden predecir el futuro, sí pueden estimar la probabilidad de que algo suceda basándose en los datos ya existentes.[412]

Ello, por un lado, puede resultar de gran utilidad tanto para los distintos Cuerpos y Fuerzas de Seguridad como para los fiscales y los jueces de instrucción, que tienen que tomar decisiones en muchas ocasiones de gran impacto en los derechos fundamentales de los ciudadandos basándose en posibles comportamientos futuros muy difíciles de precedir y, por su puesto, de advidinar, entre otros: destinar o no más vigilancia y patrullas a ciertas zonas, acordar o no una medida cautelar de privación de libertad, imponer o no una orden de protección a favor de una víctima de violencia de género, conceder o no un permiso penitenciario, o imponer o no una medida cautelar de fianza a una empresa investigada.

Hasta ahora, tanto la policía como los jueces y los fiscales han basado sus decisiones en las máximas de la experiencia, sacando conclusiones a partir de las probabilidades deducidas del agregado de los datos y de las circunstancias concretas de cada caso; inevitablemente, de los resultados de casos anteriores eminentemente iguales o muy similares; y, en ocasiones, de guías de criterios proporcionadas por la ley, por la jurisprudencia o por protocolos de actuación creados por los distintos cuerpos actuantes. Ello, no obstante, deja una toma de decisiones con gran incidencia en los derechos de los ciudadanos al criterio de los mandos policiales y de las autoridades fiscales y judiciales que, si bien han recibido extensa y completa formación para ello y cuentan con criterios legal y jurisprudencialmente establecidos, en ocasiones se ven sobrepasados e inundados de dudas, habida cuenta de que los casos concretos siempre presentan matices, y de que la predicción de comportamientos futuros resulta siempre complicada -especialmente para aquellos que no tienen apenas experiencia-. Afortunadamente, no obstante, en nuestro ordenamiento jurídico (en el ámbito judicial, que no en el policial) existe el derecho a la segunda instancia y, por ende, todas esas decisiones tomadas por los órganos judiciales pueden ser revisadas por órganos superiores que volverán a evaluar las circunstancias del caso concreto y que, según máximas de una mayor experiencia, con pleno apoyo en la ley y en la jurisprudencia, confirmarán o revocarán las resoluciones anteriores.

[412] Waldman, 2019, pág. 5.

No obstante, el uso de una tecnología como la IA para asistir o, incluso, sustituir a los humanos en la toma de decisiones policiales, fiscales o judiciales basadas en la predicción de comportamientos futuros y en la detección de posibles riesgos, tiene, sin duda, luces y sombras que conviene analizar con más profundidad, siempre sin perder de vista que el objetivo del uso de tal tecnología, tal y como se ha ido poniendo de manifiesto a lo largo del presente trabajo, es conseguir beneficiar a los humanos con las innumerables oportunidades que esta brinda, teniendo presente que la prioridad es lograr mejorar nuestro bienestar sin vulnerar ninguno de nuestros derechos ni causarnos mayores inconvenientes.

Siendo que un buen número de las herramientas que se analizarán son empleadas para la toma de decisiones en el ámbito policial (y, asimismo, en el militar y en el de los servicios de inteligencia, siendo que la línea divisoria entre tales cuerpos de seguridad resulta cada vez más difuminada, como se verá), repercutiendo luego, no obstante, en la fase de instrucción judicial (aunque en mayor o menor medida, según la herramienta de que se trate), procede hacer una distinción entre los sistemas de IA de predicción y evaluación de riesgos propios de los Cuerpos y Fuerzas de Seguridad del Estado y los de las autoridades judiciales (y, en su caso, fiscales). En relación con las primeras, no obstante, procede advertir que su análisis no se llevará a cabo de forma tan profunda como en las segundas, habida cuenta de que, si bien su estudio es sin duda muy interesante y útil, lo cierto es que se aleja un poco del concreto ámbito de la presente tesis doctoral, como consecuencia de su carácter eminentemente preventivo y policial, con menos incidencia (al menos, directa) en la instrucción judicial.

3.2.1.2. Ámbito policial. Sistemas de policía predictiva.

Como cuestión previa, procede poner de manifiesto que en los últimos tiempos, especialmente a raíz de los atentados terroristas ocurridos el 11 de septiembre del 2001 en distintas ciudades de EEUU y la posterior amenaza yihadista global, las fuerzas militares, de inteligencia y policiales de los distintos países del mundo, con la Estrategia Nacional de Seguridad publicada en 2002 por dicho país[413] como referencia, han llevado a cabo una

[413] Véase Gobierno de Estados Unidos, 2002.

gran labor de interacción y colaboración, no solo a nivel transnacional sino también nacional.

Y es que si bien, tradicionalmente, las Fuerzas Armadas y los cuerpos de inteligencia de los Estados, junto con el Ministerio del Interior, se ocupaban de definir las políticas de defensa, y las autoridades locales asumían la labor de facilitar la acción de los cuerpos policiales para poder seguir la estrategia de prevención de delitos fijada de modo central, a partir de los antedichos eventos se han ido difuminando cada vez más las líneas entre tales cuerpos de seguridad, especialmente en entornos urbanos, siendo que las grandes ciudades se han convertido en espacios donde los límites entre la seguridad nacional y la seguridad local acaban confluyendo. Así, los dispositivos que antes solo se usaban contra amenazas militares extranjeras, ahora se están implementando en las principales urbes y, entre otros, por ejemplo, las cámaras de circuito cerrado y los sistemas de identificación biométrica son herramientas que hoy en día se utilizan tanto para combatir el terrorismo, como para prevenir robos y otros delitos menores.[414]

Dicho esto, procede entrar en materia.

Ya en 1982, el politólogo James Q. Wilson y el criminólogo George Kelling publicaron la conocida "teoría de la ventana rota", que puso de manifiesto que el desorden urbano en forma de "abandono" de los espacios públicos atrae conductas antisociales y delictivas y sugirieron que, con el fin de reducir la violencia, los cuerpos policiales identificaran aquellos factores que a menudo se convertían en desencadenantes de grandes problemas e intervinieran antes de que ocurrieran los eventos dañosos, abriendo así paso a la denominada "acción preventiva" de la policía.

En los últimos años, un gran número de cuerpos policiales de todo el mundo (especialmente de EEUU) ha anunciado el uso, en su día a día de trabajo, de sistemas de IA que emplean datos estadísticos históricos para asistirles en su toma de decisiones, lo cual se conoce como el fenómeno del "*predictive policing*" o policía predictiva, definido por Martin Degeling, investigador de la Universidad de Bochum (Alemania), y Bettina Berendt, profesora de la

[414] Véase Van der Sloot, Broeders & Schrijvers, 2016, pág. 92.

Universidad Técnica de Berlín (Alemania), como aquella "*variedad de técnicas utilizadas por los cuerpos policiales para generar probabilidades delictivas, a menudo denominadas predicciones, y actuar en consecuencia*".[415]

El principal objetivo de tales sistemas es, sin duda, la optimización de recursos y el incremento de la eficacia y la eficiencia policial en la tarea de la prevención de delitos. En virtud de ello, lo que tales herramientas de IA hacen es analizar los datos históricos estadísticos que constan en las bases de datos policiales para predecir en qué áreas geográficas hay una mayor probabilidad de actividad criminal, qué perfiles de personas tienen mayores posibilidades de delinquir en el futuro, o qué tipo de gente cuenta con mayor predisposición a ser víctima de un delito, entre otros, lo que aporta una información valiosísima para los cuerpos de seguridad, que pueden así incrementar la vigilancia en aquellas zonas calificadas como "calientes" y sobre aquellos perfiles de personas calificadas con "mayor riesgo" o "mayor vulnerabilidad".

Ello, lo que consigue, desde luego, es reforzar y fomentar más que nunca el enfoque proactivo del trabajo en el ámbito policial. Y es que, a pesar de que siempre han existido tareas de vigilancia, la labor de la policía ha tendido a ser eminentemente reactiva, habida cuenta de que los agentes suelen responder a llamadas de auxilio para actuar o a informaciones de posibles comisiones de delitos para empezar a investigar, siendo que los recursos de que disponen son limitados.

En relación con lo anterior, y a modo ilustrativo, la Agencia de los Derechos Fundamentales de la Unión Europea elaboró en el año 2019 un cuadro comparativo entre los métodos de policía tradicional y los de policía predictiva que puede resultar interesante:

	ACTIVIDAD POLICIAL TRADICIONAL	**POLICÍA PREDICTIVA**

[415] Degeling & Berendt, 2018, pág. 348.

Contexto	Comisión de un delito o presentación de alerta sobre una persona en particular	Ni se ha cometido ningún delito ni se ha presentado ninguna alerta sobre una persona en particular
Aproximación	Reactiva	Proactiva
Objetivo	Detener al/los/las sospechoso/s/as	Prever dónde y cuándo pueden cometerse delitos o por/contra quién
Datos utilizados	Información específica relacionada con el caso	Información genérica relativa a varios casos
Tipo de proceso	Los procesos basados en datos y los procesos humanos se combinan	Se centra principalmente en procesos basados en grandes cantidades de datos

416

Con la finalidad de entender un poco mejor el funcionamiento y la utilidad de los sistemas de vigilancia predictiva policial a los que se está haciendo referencia, resulta conveniente poner de manifiesto que estos pueden ser divididos en cuatro grandes categorías:

1)*Métodos para predecir delitos*, que tienen como objetivo pronosticar en qué lugares y en qué tiempos (zonas geográficas, franjas horarias, días de la semana, estaciones del año, etc) se presenta un mayor riesgo de comisión de delitos;

2)*Métodos para predecir identidades delictivas*, que tienen por misión la creación de perfiles delictivos, con carácter general, para identificar a los posibles delincuentes del futuro (lo que suele estar relacionado con la existencia de antecedentes penales y pasados delictivos activos);

[416] Véase UE, 2018.

3)*Métodos para predecir víctimas de delitos*, que tienen como propósito la identificación de aquellos individuos o grupos de individuos que es más probable que resulten víctimas de un delito;[417] y

4)*Métodos para predecir delincuentes*, que tienen como finalidad identificar el riesgo concreto de que un determinado individuo delinca en el futuro.

Es importante poner de manifiesto, no obstante, que tales sistemas de IA no tienen como misión sustituir o reemplazar la vigilancia policial tradicional, sino complementarla y mejorarla, ya que cuentan con capacidad para analizar cantidades ingentes de datos (muchos más de los que podría humanamente examinar el agente con más experiencia) y extraer patrones de comportamiento para ayudar así a los Cuerpos de Seguridad, con recursos limitados, a entender y a atacar los eventuales focos de problemas, pudiendo tales resultados, desde luego, combinarse con las conclusiones y predicciones realizadas por los equipos humanos de inteligencia e investigación ya existentes en los distintos cuerpos policiales. Así, los agentes pueden diseñar estrategias e idear tácticas para prevenir o mitigar daños futuros con una mayor y más exacta -aunque, por desgracia, no siempre- información, que puede basarse en variables tales como lugares, personas/grupos o tipos de delitos.

En relación con lo anterior, procede poner de relieve que tales sistemas suelen emplearse por los cuerpos policiales para llevar a cabo, principalmente, dos tipos de funciones: la toma de decisiones estratégicas y la toma de decisiones particulares.

Así, por un lado, la toma de decisiones estratégicas (la función más extendida) se basa en la realización de predicciones agregadas o generales sobre la actividad delictiva futura, poniendo especial atención al plano geoespacial -el dónde y el cuándo pueden producirse delitos- basándose en el análisis de grandes cantidades de datos. Un ejemplo de sistema que resulta utilizado para llevar a cabo este tipo de funciones es PredPol, al que luego se hará especial referencia. Por otro lado, la toma de decisiones particulares se basa en la realización de predicciones sobre individuos o grupos concretos, delincuentes reales o

[417] Véase Perry, McInnis, Price, Smith & Hollywood, 2013.

potenciales, centrándose en identificar aquellos comportamientos que muestren mayores probabilidades de que estos cometan un delito futuro o sean propensos a desarrollar comportamientos delictivos. Un ejemplo de sistema que resulta utilizado para llevar a cabo este tipo de funciones, por su parte, es HART, al que también se hará especial referencia con posterioridad.

No cabe duda, pues, de que la actuación policial actualmente está en gran parte basada en la aplicación de sistemas y estrategias inteligentes (en parte, de IA), lo que hace que su *modus operandi* sea cada vez más parecido al de los servicios de inteligencia.

Está claro que las herramientas descritas pueden aportar múltiples ventajas y beneficios a los cuerpos policiales en el ámbito de la prevención delictiva, lo que sin duda puede resultar de gran ayuda para rebajar los niveles delictivos y hacer disminuir así, proporcionalmente, el ingente volumen de casos que llega a los juzgados de instrucción, que están altamente colapsados.

Así, actualmente, existen diversas iniciativas orientadas a dotar a las Fuerzas de Seguridad de más y mejores herramientas de IA para minimizar la actividad delictiva, tales como el denominado proyecto COPKIT, coordinado por Isdefe[418] y financiado por la Unión Europea -con una duración de treinta y seis meses (de 2018 a 2021)-, que tiene como objetivo desarrollar, con base en el marco legal vigente, aquellas herramientas de IA que empleará la policía del futuro, sin vulnerar los principios de libertad, igualdad y justicia, centrándose principalmente en el objetivo de analizar, investigar, mitigar y prevenir el uso de las nuevas tecnologías de la información y la comunicación por parte del crimen organizado y los grupos terroristas.[419]

No obstante, "no es oro todo lo que reluce", y es que la tecnología que subyace en dichos sistemas policiales no es otra que la IA y, tal y como ya se ha venido poniendo de manifiesto en la presente tesis doctoral, esta entraña algunos desafíos y riesgos que se deben afrontar, debiendo poner especialmente el foco en los que tienen impacto jurídico.

[418] Una empresa estatal propiedad del Ministerio de Defensa español.
[419] Véase COPKIT, 2018.

Así, a modo de ejemplo, en 2010, John A. Eterno, un oficial de la Policía de Nueva York retirado, profesor de Molloy College (Nueva York, EEUU), y Eli B. Silverman, profesor emérito del John Jay College of Criminal Justice and Graduate Center de la Universidad de Nueva York (Nueva York, EEUU) que había trabajado en el Departamento de Justicia de EEUU y en la Academia Nacional de Administración Pública de tal país, publicaron un estudio[420] que analizaba si los oficiales del Departamento de Policía de la ciudad de Nueva York (NYPD) -que introdujo la herramienta CompStat en 1994 y desde entonces reportó significativas disminuciones en los índices delictivos (más del 76%)- percibían que el uso de tal sistema aumentaba la presión sobre ellos para reducir las tasas de delincuencia y, de ser así, si esa presión se traducía en una contribución a la creación de estadísticas de delincuencia inexactas o poco éticas, lo que, desafortunadamente, arrojó un resultado afirmativo, a la vista de las investigaciones efectuadas.

Y es que está claro que este tipo de sistemas entrañan riesgos comunes al resto de herramientas que emplean IA (tales como la posible mala calidad de los datos y el consiguiente trato discriminatorio, la eventual falta de transparencia o la posible vulneración de derechos fundamentales en juego), si bien también cuentan con riesgos específicos que conviene enumerar.

Por un lado, respecto de los mencionados riesgos generales, procede poner de manifiesto que el mero hecho de señalar áreas, grupos de personas o incluso perfiles concretos como puntos de "riesgo", y adoptar mayores y más intensas medidas de vigilancia o control en consecuencia, presenta un dilema ético-jurídico complejo: el de la seguridad frente a la privacidad, la libertad, la presunción de inocencia y la igualdad.

En tal sentido en el congreso organizado por Amnistía Internacional los días 20 y 21 de Mayo del 2019 en Países Bajos "*PHRP expert meeting on predictive policing*", con la asistencia de expertos de diferentes áreas (policía, criminología, científicos de datos, investigadores académicos y sociedad civil), se llegó a la conclusión de que los datos empleados en los sistemas analizados nunca reflejan la realidad completa, sino solo una parte de ella, y "*en particular, en muchos casos se refleja el enfoque y las prioridades anteriores de la policía con respecto a ciertas áreas, grupos o tipos específicos de delitos,*

[420] Véase Eterno & Silverman, 2010, págs. 426-449.

enfoques que a menudo pueden estar afectados por un sesgo estructural." No obstante, se proyectó una posible solución, la de "*cambiar la forma en que se usa la información del algoritmo: alejarse de la concepción de herramienta de toma de decisiones (dónde y a quién vigilar) para usarla como herramienta de diagnóstico: ¿por qué la información es la que es? P.ej. si un determinado grupo de personas o una determinada zona muestra una mayor prevalencia de la delincuencia, ¿cuál es la razón que hay detrás de eso? Ya que puede ser un indicador de un enfoque policial sesgado que se centra excesivamente en ese grupo o área. Esto permitiría a la policía reflexionar críticamente sobre su relación con esa comunidad y sobre cómo podrían mejorar su enfoque. También podría ayudar a profundizar en las causas de la delincuencia (por ejemplo, pobreza, falta de oportunidades, exclusión social) y abordarlas en lugar de optar por un enfoque meramente policial*."[421]

Asimismo, respecto del derecho a la presunción de inocencia, considero conveniente traer a colación la, cuanto menos sorprendente, tendencia que ha adoptado en los últimos tiempos el Tribunal Supremo de EEUU. Y es que, si bien en virtud de lo dispuesto en la 4ª Enmienda de la Constitucion de tal país (que prohíbe intervenciones policiales injustificadas y establece los requisitos necesarios para dictar órdenes de busca, entrada y registro, etc) resulta necesario que todos aquellos registros e intervenciones policiales que no sean rutinarios estén justificados por la previa existencia de una "sospecha razonable" individualizada (no genérica o abstracta) de que se está cometiendo o se va a cometer un hecho delictivo[422], en los últimos años el Alto Tribunal decidió rebajar los estándares requeridos para considerar constitucionales las intervenciones policiales en aquellas áreas consideradas de alto riesgo (que, no obstante, no han sido definidas de forma detallada), lo que sin duda deja abierta la puerta al libre empleo de los sistemas de policía predictiva analizados[423].

En relación con lo expuesto, en el ámbito de la Unión Europea, la directora ejecutiva del AI Now Institute de la Universidad de Nueva York (Nueva York, EEUU), Andrea Nill

[421] Amnistía Internacional, 2019, pág. 2.
[422] Véase Estados Unidos vs Montoya de Hernandez, 473 U.S. 531, 538 1985.
[423] En concreto, en el caso Illinois v. Wardlow Illinois vs Wardlow, 528 U.S. 119, 124, 2000, el Tribunal Supremo de EEUU sostuvo que la fuga de un sospechoso, en caso de intervención policial, en un área con alto índice de criminalidad, ya de por sí implicaba una sospecha razonable bastante para poder llevar a cabo una detención y/o un registro.

Sánchez, en la Comisión LIBE del Parlamento Europeo que se celebró el 18 de febrero del 2020 y versó sobre "La IA en el Derecho Penal y su uso por parte de las autoridades policiales y judiciales en materia penal", y la Vicepresidenta de Política Digital de la Comisión Europea, Margrethe Vestager, en su discurso pronunciado en el European AI Forum el 30 de junio del mismo año, ya advirtieron de la necesidad de que el uso de las tecnologías de policía predictiva dentro de sus fronteras quedara estrictamente limitado por el respeto a los derechos fundamentales en juego.

Por otro lado, respecto de los riesgos específicos, en primer lugar procede hacer referencia al peligro de que, con su uso, los Cuerpos y Fuerzas de Seguridad se queden en la superficie de la solución y no ahonden en el origen del problema, ya que es muy tentador (habida cuenta de la escasez de recursos humanos y materiales que acecha a muchos cuerpos policiales), simplemente seguir lo que dicta el sistema algorítmico sin plantearse mucho más. Ello, no obstante, opera como mero "parche", ya que si bien, seguramente, servirá para aumentar los niveles de eficacia de las operaciones policiales, puesto que incrementará el número de detenciones e incluso el de frustración previa de actos delictivos, lo cierto es que lo necesario para realmente prevenir la actividad delictiva a medio y largo plazo es entender el origen del problema y tomar medidas para solucionarlo. Así, ante la pregunta: "¿por qué hay más riesgo en una zona determinada o en un perfil concreto?", la simple respuesta: "porque lo dice el sistema de IA", es muy peligrosa.

En segundo lugar, en ocasiones los sistemas de IA no aportan una utilidad táctica demasiado alta, habida cuenta de que marcan como puntos calientes zonas geográficas de grandes dimensiones, o establecen perfiles delictivos demasiado genéricos, lo que hace muy difícil adoptar medidas concretas que permitan obtener resultados de éxito elevado.

En tercer lugar, en muchos casos un gran riesgo de los sistemas de IA analizados es la ausencia de realización de evaluaciones posteriores por parte de los cuerpos policiales que los emplean. Y es que, la efectividad de los sistemas de vigilancia predictiva depende, en gran medida, de su posterior comprobación o análisis, siendo que ello resulta fundamental para averiguar así si las predicciones realizadas y las intervenciones llevadas a cabo como respuesta por los cuerpos policiales han tenido un verdadero impacto en los datos delictivos o no, lo cual deviene imprescindible para, entre otras cosas, actualizar información y, sobre

todo, planear futuras intervenciones identificando áreas de mejora, solventando errores y distribuyendo recursos.

Y, finalmente, otro desafío específico que entrañan este tipo de sistemas es su escasa efectividad en la predicción de determinados delitos, tal y como asegura la profesora de la Universidad de Nueva Gales del Sur ubicada en Sidney (Australia) Lyria Bennett Moses, Directora del Allens Hub for Technology, Law and Innovation, que manifiesta: "*Al observar delitos como el robo en vivienda habitada, se puede crear un modelo predictivo bastante útil porque algunas áreas tienen tasas más altas que otras y existen patrones; sin embargo, funciona muy mal para el secuestro o la violencia doméstica, en este último caso porque gran parte del delito no se denuncia*." [424]

Como consecuencia de los múltiples riesgos que este tipo de sistemas entrañan, la LO 7/21, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, establece ciertas limitaciones al tratamiento de los datos personales lo que, por ende, afecta al uso de esta clase de herramientas con el objetivo de evitar que se provoquen vulneraciones de derechos fundamentales cuando son empleadas por las autoridades, que en virtud de lo dispuesto en el artículo 4 de tal cuerpo legal son consideradas como: "*toda autoridad pública que tenga competencias encomendadas legalmente para el tratamiento de datos personales con alguno de los fines previstos en el artículo 1*.", y en particular:

a) Las Fuerzas y Cuerpos de Seguridad.
b) Las Administraciones Penitenciarias.
c) La Dirección Adjunta de Vigilancia Aduanera de la Agencia Estatal de Administración Tributaria.
d) El Servicio Ejecutivo de la Comisión de Prevención del Blanqueo de Capitales e Infracciones Monetarias.
e) La Comisión de Vigilancia de Actividades de Financiación del Terrorismo.
f) las Autoridades judiciales del orden jurisdiccional penal y el Ministerio Fiscal.

[424] Bennett Moses, 2020.

En relación con ello, es importante poner de manifiesto que por tratamiento de datos se entiende "*cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción*", de acuerdo con lo dispuesto en el artículo 5.b) de la LO 7/21, de 26 de mayo (en los mismos términos que los previstos en el artículo 4.2 RGPD).

Asimismo, es importante remarcar que la mayor parte de las herramientas objeto de análisis en la presente Sección tratan datos personales, por lo que les resulta de aplicación lo dispuesto en los artículos 6 a 12 del mencionado cuerpo legal (en lo relativo a los principios generales de tratamiento, deberes de colaboración, plazos de conservación y revisión, distinción entre categorías de interesados, verificación de calidad y licitud de tratamiento), que serán desarrollados en más profundidad cuando se haga referencia a las herramientas de IA de investigación penal que emplean datos biométricos[425]; en especial lo previsto en los artículos 13 y 14; y, por supuesto, lo previsto en el Capítulo III respecto de los Derechos de las personas, en el Capítulo IV respecto de los responsables y los encargados del tratamiento, en el Capítulo V respecto de las transferencias de datos personales a terceros países que no sean miembros de la UE o a organizaciones internacionales, en el Capítulo VI respecto de las Autoridades de Protección de Datos Independientes, en el Capítulo VII respecto de las reclamaciones, y en el Capítulo VIII respecto del régimen sancionador.

En concreto, además, en su artículo 5.d) la mencionada LO (así como el artículo 4.4 del RGPD) define lo que es la "elaboración de perfiles" del modo siguiente: "*toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física*".

[425] Véanse págs. 416-436.

De acuerdo con lo expuesto, y en especial, en virtud de lo previsto en el artículo 14 del antedicho cuerpo legal, con carácter general en el ámbito de la UE (ya que esta LO no es más que la transposición de la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016) queda prohibido el uso de las herramientas que estamos analizando en caso de que conlleven decisiones basadas únicamente en un tratamiento automatizado de datos personales, incluida la elaboración de perfiles, siendo que su utilización produce efectos jurídicos que afectan significativamente a los ciudadanos, salvo que ellose autorice expresamente por una norma con rango de ley o por el Derecho de la UE. Y es que el legislador deja libertad a la UE y a los Estados Miembros para que permitan, de forma expresa, el uso de este tipo de sistemas mediante normas que deberán contar con ciertas garantías, a saber: la previsión de medidas adecuadas para salvaguardar los derechos y libertades de los interesados, incluyendo el derecho a obtener la intervención humana en el proceso de revisión de la decisión adoptada. En España, no obstante, todavía no contamos con ninguna norma habilitante en tal sentido, por lo que no pueden ser empleadas tales herramientas en los antedichos términos.

Y la analizada LO, además, va un paso más allá: prohíbe con carácter general que las decisiones basadas únicamente en un tratamiento automatizado, incluida la elaboración de perfiles, se apoyen en categorías especiales de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas o la afiliación sindical, datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o a la vida sexual o a la orientación sexual de una persona física. No obstante, también se prevé como excepción a esta prohibición la adopción de las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, lo cual deberá ser determinado por el legislador en cada momento.

Y, finalmente, la mencionada LO establece una prohibición estricta, sin condiciones, aplicable al uso de este tipo de herramientas: no se permite la elaboración de perfiles que dé lugar a una discriminación de las personas físicas sobre la base de las categorías especiales de datos personales antedichas. En mi opinión, no obstante, tal interdicción expresa resulta obvia, habida cuenta de que ya la Carta de Derechos Fundamentales de la UE y la propia Constitución Española prohíben todo tipo de discriminación, si bien nunca

está de más recordar la necesidad de respetar los derechos fundamentales, en especial en relación con este tipo herramientas.

Dicho lo anterior, se deduce, pues, que las herramientas objeto de análisis estarán permitidas en España únicamente en dos casos: siempre y cuando su uso esté previsto legalmente en los términos del artículo 14 (lo cual hoy en día no ocurre); y siempre y cuando no conlleven decisiones basadas únicamente en un tratamiento automatizado de datos personales, incluida la elaboración de perfiles, es decir: siempre que haya además una intervención humana cualificada. Y es que en relación con esto último, procede poner de manifiesto que no es lo mismo el hecho de que la policía detenga a una persona simplemente porque una herramienta de IA de evaluación de riesgos ha determinado que es peligrosa y tiene altas probabilidades de reincidir, que el hecho de que la policía tome esa decisión basándose no solo en el resultado del sistema sino también en otras circunstancias y elementos periféricos que, debidamente valorados, le lleven a la misma conclusión.

Respecto de este segundo supuesto (a saber, el de las herramientas policiales de evaluación de riesgos que no conlleven decisiones basadas únicamente en un tratamiento automatizado de datos personales, incluida la elaboración de perfiles), debe hacerse una importante precisión, no obstante, para el supuesto de que se traten categorías especiales de datos personales.

Y es que el artículo 13.1 de la mencionada LO 7/21, de 26 de mayo, es claro al afirmar que:

"*El tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas o la afiliación sindical, así como el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, los datos relativos a la salud o a la vida sexual o a la orientación sexual de una persona física, sólo se permitirá cuando sea estrictamente necesario, con sujeción a las garantías adecuadas para los derechos y libertades del interesado y cuando se cumplan alguna de las siguientes circunstancias:*

a) Se encuentre previsto por una norma con rango de ley o por el Derecho de la Unión Europea.

b) Resulte necesario para proteger los intereses vitales, así como los derechos y libertades fundamentales del interesado o de otra persona física.

c) Dicho tratamiento se refiera a datos que el interesado haya hecho manifiestamente públicos."

De acuerdo con ello, pues, aquellas herramientas de evaluación de riesgos que traten datos personales calificados como especiales solo resultarán permitidas cuando su tratamiento resulte estrictamente necesario, con sujeción a las garantías adecuadas para los derechos y libertades del interesado (lo cual, no obstante, puede dar lugar a interpretaciones subjetivas) y siempre y cuando: o bien se halle previsto por una norma con rango de ley o por el Derecho de la Unión Europea (lo que no ha ocurrido todavía), o bien resulte necesario para proteger los intereses vitales, así como los derechos y libertades fundamentales del interesado o de otra persona física (lo que asimismo puede dar lugar a interpretaciones subjetivas) o bien se refiera a datos que el interesado haya hecho manifiestamente públicos (lo que se entiende como un consentimiento, pero también puede dar lugar a discrepancias interpretativas), debiendo pues atenerse a cada caso concreto para determinar la legitimidad del uso de las distintas herramientas objeto de análisis.

Visto lo anterior, bajo mi punto de vista, desde luego considero que hace falta una regulación específica y clara que haga referencia expresa al uso de este tipo de herramientas, tal y como la propia LO requiere, puesto que el terreno que se pisa en la actualidad es ambiguo y puede dar lugar a malas interpretaciones y malas prácticas habida cuenta del vacío legal que existe.

En relación con ello, no obstante, procede anunciar que más pronto que tarde puede haber buenas noticias, puesto que la ya mencionada Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, regula de forma específica gran parte de las mencionadas herramientas, calificándolas como sistemas de alto riesgo en su Anexo III (en relación con

lo dispuesto en su artículo 6.2) y sometiendo su uso, por ende, a unos requisitos muy estrictos.[426]

Y es que en el mencionado Anexo III, en concreto, se hace referencia a:

"*-Asuntos relacionados con la aplicación de la ley:*

a) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para llevar a cabo evaluaciones de riesgos individuales de personas físicas con el objetivo de determinar el riesgo de quecometan infracciones penales o reincidan en su comisión, así como el riesgo para las potenciales víctimas de delitos; (…)

e) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para predecir la frecuencia o reiteración de una infracción penal real o potencial con base en la elaboración de perfiles de personas físicas, de conformidad con lo dispuesto en el artículo 3, apartado 4, de la Directiva (UE) 2016/680, o en la evaluación de rasgos y características de la personalidad o conductas delictivas pasadas de personas físicas o grupos;

f) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para la elaboración de perfiles de personas físicas, de conformidad con lo dispuesto en el artículo 3, apartado 4, de la Directiva (UE) 2016/680, durante la detección, la investigación o el enjuiciamiento de infracciones penales;

g) sistemas de IA destinados a utilizarse para llevar a cabo análisis sobre infracciones penales en relación con personas físicas que permitan a las autoridades encargadas de la aplicación de la ley examinar grandes conjuntos de datos complejos vinculados y no vinculados, disponibles en diferentes fuentes o formatos, para detectar modelos desconocidos o descubrir relaciones ocultas en los datos."

Sentadas las anteriores bases, pues, procede entrar a examinar cada uno de los distintos tipos de sistemas de vigilancia policial predictiva que existen en la actualidad y sus posibles implicaciones jurídicas.

[426] Véanse págs. 192-197.

a.1) Sistemas de mapeo predictivo ("*Predictive mapping*" o "*Place-oriented techniques*")

Está claro, siendo que así se desprende de multitud de estudios de criminología elaborados a lo largo de los tiempos, que es posible la predicción del crimen y, por ende, la prevención del mismo, habida cuenta de que los delincuentes tienden a cometer delitos en su denominada "zona de confort" (geotemporal), que en la mayoría de ocasiones es aquella en que han podido cometer en el pasado un delito similar con éxito. Según Jeff Brantingham, antropólogo de la Universidad estadounidense de California-Los Angeles (UCLA) especialista en sistemas de vigilancia predictiva: "*Los detractores quieren hacernos creer que los humanos son demasiado complejos y aleatorios*" (…) "*Pero los humanos no son tan aleatorios como pensamos*" (…) "*En cierto sentido, el crimen es solo un proceso físico, y si puedes explicar cómo se mueven los delincuentes y cómo se mezclan con sus víctimas, puedes entender muchísimo*".[427]

Y es que las estadísticas pueden dar, a un nivel muy preciso y exacto, información valiosísima sobre los "puntos calientes" ("*hotspots*") de criminalidad con la que se pueden ir construyendo algoritmos para elaborar modelos de policía predictiva cada vez más efectivos y sofisticados. Así, por ejemplo, se sabe que en Bogotá (Colombia), en el ámbito temporal, el pico máximo de violencia homicida en el periodo 2012-2013 tuvo lugar los domingos de madrugada, entre las 2:30 y las 3:30h; y, en el ámbito espacial, en Cali (Colombia), el 50% de todos los homicidios, en un periodo semanal, tuvo lugar en el 1,28% de los segmentos de vía existentes en dicha ciudad[428]. Y tales datos, agregados a gran escala, pueden confirgurar un sistema de *big data* que sin duda aporta un valor sin precedentes a la lucha contra el crimen.

No hay duda de que el origen del denominado modelo de *predictive mapping* o mapeo predictivo, que empezó a sonar con fuerza en el ámbito policial a partir del año 2008, tiene un nombre propio: el del comisario de policía estadounidense William J.Bratton, que de forma conjunta con la policía de Los Ángeles (Los Angeles Police Department -LAPD-),

[427] Rubin, 2010.
[428] Véase Mejia, Ortega, & Ortiz, 2015.

trabajó y fomentó la creación y la utilización de sistemas de IA para mejorar los datos delictivos de dicha ciudad.

En relación con ello, y con el objetivo de dar a conocer tal nueva metodología y abrir un debate sobre la misma entre los distintos agentes implicados y potencialmente interesados (que eran muchos, especialmente tras los atentados ocurridos el 11 de septiembre de 2001 en EEUU, en que se puso de manifiesto la necesidad de introducir cambios para incrementar la efectividad policial, especialmente en el ámbito de la predicción y prevención de delitos), el National Institute of Justice (NIJ) de tal país decidió celebrar dos simposios sobre los nuevos métodos de vigilancia predictiva ("*Predictive Policing Symposiums*"), que tuvieron lugar los días 18 y 19 de noviembre del 2009 en Los Ángeles (California, EEUU), y el 2 y 3 de junio del 2010 en Providencia (Rhode Island, EEUU). En concreto, en el primero de estos, Kristina Rose, directora adjunta del NIJ, atribuyó al mencionado comisario William J.Bratton el mérito de haber llevado los sistemas de vigilancia predictiva a la vanguardia.[429]

Así, en 2011, el Departamento de Policía de Los Angeles (LAPD) inició un experimento para poner a prueba la efectividad del método de policía predictiva, y los resultados demostraron científicamente la eficacia del modelo. Dicho ensayo se desarrolló en la zona de Foothill (Los Angeles, EEUU), donde la división de policía encargada del área procedió a distribuir diariamente a los agentes mapas de la ciudad para que llevaran a cabo sus rutas de trabajo. La clave estaba en que, sin que los agentes tuvieran conocimiento de ello, algunos días en dichos mapas se fijaba una ruta policial basada en métodos tradicionales y, otros días, las rutas policiales se calculaban utilizando el algoritmo que proporcionaba el sistema de policía predictiva. Tras ello, los resultados fueron alentadores, ya que, para sorpresa de muchos, el sistema basado en algoritmos proporcionó el doble de precisión que las prácticas tradicionales, habiendo llegado a disminuir un 12% los delitos contra la propiedad los días en que se utilizaron los mapas basados en tal método.[430] [431]

[429] Véase Rose, 2009.
[430] Véase Mohler & otros, 2014, págs. 1399-1411.
[431] Véase Wolpert, 2015.

A lo largo de los años, el interés (no solo de los cuerpos policiales de alrededor del mundo, sino también del público en general) por el uso de tales sistemas ha ido proliferando. No obstante, tales avances han estado especialmente marcados por la aparición de diversas noticias relativas a las bonanzas y los riesgos de los sistemas empleados, a los que luego se hará especial mención.

En la actualidad, numerosos cuerpos policiales de EEUU (entre otros, Los Angeles, Nueva York y Memphis), América Latina (entre otros, Chile[432], y previsiblemente, en un futuro próximo, Colombia[433]), Europa (entre otros Inglaterra, Alemania e Italia) o Asia (entre otros, Singapur[434]) se apoyan en estos métodos de vigilancia predictiva basados en el uso de IA, y no es poca la literatura existente sobre ellos.

No obstante, respecto de América Latina, los expertos locales advierten de que: "*Nuestros gobiernos, bien o mal intencionados, y frecuentemente desprovistos de conocimiento sobre los alcances de las nuevas tecnologías, tienden a considerar la adquisición de herramientas tecnológicas policiales como la solución automática a problemas multidimensionales. Esto es un grave error. Generalmente se deja de lado, así, el análisis del perfil y la ética necesaria en el personal que opera la tecnología, de las condiciones organizativas que rodean su introducción, del impacto de las nuevas tecnologías en la protección de los derechos humanos, de su impacto en los niveles de corrupción y en la concentración de poder, del balance entre el costo económico de adquirir la nueva tecnología y su impacto en el logro de una seguridad efectiva*."[435], lo cual podría extenderse a la totalidad de países que autorizan el uso de la tecnología analizada.

En relación con lo expuesto, es importante hacer una enumeración de las herramientas de IA de mapeo delictivo más empleadas por las Fuerzas y Cuerpos de Seguridad de alrededor del mundo, a pesar de que, todo hay que decirlo, no es tarea fácil obtener información al respecto habida cuenta del gran sigilo y cautela con que tratan este tema algunos cuerpos policiales, bajo el paraguas de la confidencialidad y, en mi opinión, principalmente por miedo a ser sometidos a un escrutinio público demasiado severo.

[432] Véase Vak, 2019.
[433] Véase Zuñiga, 2019.
[434] Véase Kang Wei & See Kiat, 2014.
[435] González, Casas & Mesías, 2018, pág. 22.

Por un lado, procede hacer especial mención al sistema CompStat ("*Computer Comparison Statistics*"), ya mencionado con anterioridad, que fue creado a mediados de la década de los años 90 por el entonces comisario del Departamento de Policía de la ciudad de Nueva York (NYPD) William J.Bratton y su comisario adjunto Jack Maple, habiéndose implementado por primera vez en dicho cuerpo policial como medida de ayuda para controlar el crimen y mejorar la calidad de vida de los ciudadanos, y habiendo proliferado su uso posteriormente en distintos cuerpos policiales de todo EEUU.

En concreto, el NYPD ha asegurado que gracias a la utilización del sistema CompStat se ha conseguido reducir la delincuencia en la ciudad en un 75% en aproximadamente 20 años[436], lo cual es un resultado aparentemente magnífico. Desafortunadamente, no obstante, no siempre se ha empleado tal sistema como una herramienta de ayuda sino también como un arma policial que en muchos casos ha creado una fuerte dependencia y ha traído consecuencias muy negativas, tal y como se recogió en el informe publicado por ProPublica en el año 2016.[437]

En relación con ello, Eli Silverman, criminólogo, y John Eterno, un ex oficial de distrito del NYPD, ya en su libro "*The Crime Numbers Game*"[438] pusieron de manifiesto que en la ciudad de Nueva York, el uso de CompStat fue la causa directa de diversas prácticas policiales abusivas llevadas a cabo en comunidades de color, habiendo contribuido a que ciertos oficiales de policía publicaran cifras de crimen sesgadas que daban a entender que en algunas de dichas comunidades se estaban cometiendo más delitos que en otras. En concreto, en 2010, Sharif Stinson, un residente del East Village de la mencionada ciudad, presentó una demanda colectiva contra el Departamento de Policía de tal ciudad argumentando con éxito que este emitió más de 900.000 citaciones criminales ilegales[439].

Por otro lado, procede hacer alusión a PredPol, un sistema de vigilancia predictiva que ha sido sometido a más de un millón de horas de pruebas oficiales en departamentos de todos los tamaños en todo el mundo y que se basó en un proyecto de investigación llevado a cabo

[436] Véase McHugh, Stulberger & Dienst, 2016.
[437] Larson, Mattu, Kirchner & Angwin, 2016.
[438] Eterno & Silverman, 2012.
[439] Véase Weiser, 2017.

entre el Departamento de Policía de Los Ángeles (LAPD) y la Universidad estadounidense de California-Los Angeles (UCLA). En relación con ello, Jeff Brantingham, profesor de antropología de tal universidad, adaptó los trabajos que había realizado para pronosticar víctimas en el campo de batalla de Iraq, financiados por el Pentágono, para predecir el crimen en el ámbito policial, habiendo patentado su investigación, que culminó con la fundación de una compañía con ánimo de lucro llamada PredPol, LLC., que comercializó un sistema con dicho nombre capaz de proporcionar, con éxito, predicciones en tres ámbitos: tipo de delito, ubicación, y fecha y hora del mismo.

Tal y como se expone en su propia página web, PredPol (utilizado por más de 60 departamentos de policía en EEUU) se basa en una investigación académica detallada sobre las causas de formación de patrones de delincuencia y utiliza datos del sistema de gestión de registros "*Records Management Systems*" (RMS) de su propia agencia para obtener información delictiva actual e histórica, introduciéndola luego en su algoritmo de aprendizaje automático para crear sus predicciones, prestando especial atención a la necesidad de que los datos empleados sean precisos y completos.

Según se desprende de la mencionada página web, hay tres aspectos del comportamiento del delincuente que se introducen en el modelo matemático de PredPol:

-victimización repetida, que hace referencia a la idea de que es "racional" que los delincuentes regresen a los lugares donde han tenido éxito previamente (tal y como ya se ha avanzado con anterioridad);
-victimización casi repetida, que hace referencia a la idea de que no solo es racional que los delincuentes regresen a aquellos lugares donde tuvieron éxito previo, sino que también lo es la idea de que decidan delinquir en circunstancias/ámbitos/lugares próximos o similares; y
-búsqueda local, que hace alusión a la idea de que los delincuentes rara vez se trasladan muy lejos de sus puntos clave de actividad, tales como su hogar, trabajo y lugares de juego, lo que significa que los delitos tienden a agruparse.

En concreto, PredPol, a diferencia de otras compañías que mantienen reservados sus algoritmos, bajo el paraguas del secreto de empresa, tiene publicado (y patentado) el suyo:

$$\frac{\partial A}{\partial t} = B + \frac{\eta D}{4}\nabla^2 A - \omega A + \theta\omega\delta$$

[440]

No obstante, la policía de Kent (Reino Unido), por ejemplo, dejó de usar el antedicho *software* por entender que el valor añadido que aportaba no era del todo convincente[441], por lo que, en cualquier caso, deben ponerse en cuestión las bonanzas del sistema expuestas por la compañía que lo comercializa.

Existen, no obstante, otros *software* capaces también de llevar a cabo las tareas de predicción delictiva que realizan los ya mencionados CompStat y PredPol, como ocurre con el sistema CrimeScan, que a través de diversos indicadores (delitos menores, llamadas a emergencias, etc) tiene capacidad para predecir con hasta una semana de antelación y alta precisión, los puntos calientes (espacio-temporales) de violencia[442]; o el sistema creado por IBM empleado en Memphis (EEUU), en el seno del proyecto llamado "*Blue CRUSH*" ("*Criminal Reduction Utilizing Statistical History*"), que ha conseguido reducir el crimen en un 30% en dicho Estado[443], si bien se considera que, al ser CompStat y PredPol los pioneros, merecía la pena hacer especial referencia a ellos a título de ejemplo.

A pesar de todo lo expuesto, no obstante, conviene destacar que, a la vista de la gran cantidad de riesgos para los derechos fundamentales (especialmente para el derecho a la privacidad y a la protección de datos, a la libertad, a la presunción de inocencia, a la igualdad y a la no discriminación) que tales sistemas presentan, y a la vista asimismo del incesante incremento de la indignación popular por ciertas prácticas policiales (especialmente en EEUU), resulta patente la necesidad de llevar a cabo cambios en el sistema policial. Y tales cambios, sin duda, pasan por intentar reducir al máximo los sesgos en sus actuaciones, razón por la cual ahora más que nunca los sistemas de policía predictiva están en el foco público, ya que como ya se ha advertido, en muchas ocasiones vienen a

[440] PredPol, s.f..
[441] Schroeter, 2018, pág. 21.
[442] Neill, 2012, pág. 3.
[443] Véase Armonk, 2010.

perpetuar patrones de actuación pasados que mucho distan de lo que sería deseable en un Estado de Derecho.

Como consecuencia de ello, ya se han venido acordando medidas radicales, tales como la prohibición de utilizar este tipo de sistemas de IA. Así, la ciudad californiana de Santa Cruz (EEUU), si bien fue, hace una década, de las primeras urbes estadounidenses en emplear tal tecnología, ha sido la primera en vedarla[444], habiendo anunciado que tal restricción será permanente salvo que se publique una adecuada legislación que impida eficazmente que se perpetuen los sesgos policiales existentes.

a.2) Sistemas de identificación predictiva ("*Predictive identification systems*" o "*Person-oriented techniques*")

Una vez fijadas las bases conceptuales y analizados los potenciales riesgos y beneficios de las tecnologías de mapeo predictivo presentes en el día a día de cuerpos policiales del mundo entero, procede hacer especial mención a las denominadas herramientas de idetificación predictiva o de creación de perfiles criminales ("*predictive identification systems*"), habida cuenta de la gran cantidad de concretas implicaciones jurídicas (y éticas) que pueden tener, particularmente en el ámbito de la Unión Europea. No obstante, tal y como ya se ha adelantado anteriormente, siendo que esta tesis doctoral tiene por objeto centrarse en el uso de la IA en el ámbito de la investigación judicial de delitos, no de la prevención, no se llevará a cabo un minucioso y extenso análisis de tal tipo de sistemas, resultando fundamental, no obstante, dejar patentes las utilidades y peligros que pueden entrañar, ya que en ocasiones pueden afectar de forma directa o indirecta al trabajo de los órganos instructores.

Los mencionados sistemas emplean fundamentalmente técnicas de IA de elaboración de perfiles ("*profiling tools*") mediante las cuales se extrae una gran cantidad de datos (función denominada "minería de datos") y se procede a su análisis (función denominada "procesamiento") con el fin de hallar o trazar ciertos patrones o tipos de conducta que permitan clasificar a los ciudadanos en distintas categorías. Así, en el ámbito de la policía

[444] Véase Uberti, 2020.

predictiva, por ejemplo, ciertas personas se catalogan, en función de los datos que se han obtenido sobre ellas y su posterior análisis y procesamiento, por el nivel de riesgo que representan para la sociedad o por el nivel de probabilidad de ser víctimas de un delito.

La creación de perfiles, especialmente delictivos, puede llevarse a cabo por sistemas algorítimicos a partir de las informaciones o datos recopilados a través de distintas herramientas de IA, a saber, entre otras: dispositivos de reconocimiento facial (cámaras ubicadas en la vía pública, circuitos cerrados de televisión, etc), de imágenes (registros de matrículas, sensores de detección de actividad inusual, etc) y de voz; registros de huellas dactilares, de ADN, etc; cartografía (geolocalización); análisis o cruce de datos de dispositivos móviles, de servicios electrónicos privados y públicos (compras electrónicas, actividad en las redes sociales, correo electrónico, mensajería instantánea, etc), registros financieros, registros de nombres de pasajeros, registros de vehículos, información económica, jurídica, fiscal, etc; "Internet de las cosas" (dispositivos inteligentes y domótica, etc); herramientas de minería o análisis de textos o datos; y análisis semántico, entre otros.

Así, en el ámbito europeo, policías de distintos países emplean diversos programas de *software* de los descritos. Entre otros, cabe destacar que en el Reino Unido los cuerpos policiales de Londres, las Tierras Medias Occidentales y Avon y Somerest, cuentan con distintos programas de identificación predictiva.

Por un lado, la policía de Londres ("*Metropolitan Police*") ya en 2012 desarrolló la herramienta denominada "*The Gang Violence Matrix*" (GVM Matrix) con el fin de identificar y evaluar el riesgo que tenían los pandilleros de tal ciudad de ser delincuentes y víctimas y prevenir la comisión de delitos, especialmente los más graves. Así, el programa detecta a aquellos pandilleros más violentos, con el objetivo de estrechar el cerco policial frente a ellos, y a aquellos que han sido víctimas reiteradas de delitos y, por tanto, necesitan de más protección y apoyo policial o social para alejarse de las pandillas.[445]
No obstante, en 2018, una investigación de Amnistía Internacional[446] advirtió de la naturaleza discriminatoria del mencionado programa de identificación predictiva,

[445] Véase Metropolitan Police, s.f..
[446] Véase Amnistía Internacional, 2018.

habiéndolo definido como: "*la herramienta equivocada para el problema equivocado: un sistema racialmente discriminatorio que estigmatiza a los jóvenes negros por la música que escuchan o su comportamiento en las redes sociales*"[447], lo cual fue asimismo puesto de manifiesto ese mismo año por el Alcalde de Londres Sadiq Khan, que urgió a la policía a revisar el sistema[448] (aunque con posterioridad, en febrero de 2020 valoró positivamente los cambios introducidos[449]), y por el organismo de control de protección de datos del Reino Unido, que tras una investigación concluyó que tal herramienta vulneraba gravemente las leyes de protección de datos del país, lo que provocó que la policía londinense eliminara a casi cuatrocientas personas del sistema.[450]

Posteriormente, en mayo del 2019, la policía de Londres lanzó en el Sureste de la ciudad, bajo la dirección del Inspector Jefe Andy Briers, el programa "*Concern Hub*", resultado de una colaboración con múltiples autoridades locales, con el objetivo de identificar a aquellas personas con riesgo de involucrarse en el mundo pandillero de la capital británica e intentar reducir el número de homicidios que se producían en tal entorno.[451]

Por otro lado, la policía de las Tierras Medias Occidentales ha liderado el proyecto denominado "*National Data Analytics Solution-Most Serious Violence*" (MSV), en colaboración con otras ocho fuerzas policiales, que fue lanzado en el año 2019 con el objetivo de crear un sistema analítico avanzado (combinación de estadísticas e IA) y centralizado para toda la policía de Reino Unido, con intención de asegurar que esta contara con toda la información disponible en los distintos cuerpos policiales a través de la conexión de sus datos a fin de poder evaluar riesgos y priorizar recursos.[452]

Y, finalmente, la policía de Avon y Somerset, emplea el programa "*Qlik Sense*", que tiene un doble objetivo: predecir, por un lado, la probabilidad de que alguien cometa un determinado delito y, predecir, por otro lado, la probabilidad de que alguien se vuelva violento durante una detención.[453]

[447] Amnistía Internacional, 2020.
[448] Véase Gobierno de Reino Unido, 2018.
[449] Véase Gobierno de Reino Unido, 2020.
[450] Véase BBC, 2020.
[451] Véase Agerholm, 2018.
[452] Véase Gobierno de Reino Unido, 2019.
[453] Véase Radnor, 2017.

En el estricto ámbito de la UE, asimismo, en los últimos años, especialmente, han proliferado las iniciativas basadas en la utilización de herramientas de IA para la elaboración de perfiles a través de datos con fines de seguridad, principalmente en el ámbito antiterrorista.

Así, en Alemania, por ejemplo, la policía federal en el año 2017 desarrolló, en cooperación con expertos en psicología forense de la Universidad de Constanza (Alemania), el sistema RADAR-iTE ("*Regelbasierte Analyse potentiell destruktiver Täter zur Einschätzung des akuten Risikos – Islamistischer Terrorismus*") -con vigencia ahora de su segunda versión, RADAR-iTE 2.0.- con el fin de elaborar una lista de los terroristas yihadistas potencialmente más peligrosos. Tal listado se obtiene mediante el análisis del "comportamiento observable" de los sospechosos (más que su ideología), empleando la información que se recopila a través de las respuestas a un cuestionario de setenta y tres preguntas, que sirven al sistema para otorgar un grado de riesgo moderado, notable o alto. Además, la policía lleva a cabo una evaluación caso por caso en una segunda etapa, a través del sistema RISKANT ("análisis de riesgo de quienes se inclinan a actuar por motivos islamistas"), que se desarrolló entre 2017 y 2020. [454]

Asimismo, se han publicado en la UE diversas normativas dirigidas a dar base y cobertura legal a dicho tipo de sistemas, entre otras, la Directiva (UE) 2016/681 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la utilización de datos del registro de nombres de los pasajeros (PNR) para la prevención, detección, investigación y enjuiciamiento de los delitos de terrorismo y de la delincuencia grave, que obliga a las compañías aéreas a recopilar todos los datos personales de los pasajeros que viajan desde el territorio de la UE a terceros países y compartirlos con todos los Estados miembros con el fin de identificar determinadas categorías de "pasajeros de alto riesgo" que necesitan una mayor atención; y la Directiva (UE) 2018/843 del Parlamento Europeo y del Consejo, de 30 de mayo de 2018, por la que se modifica la Directiva (UE) 2015/849 relativa a la prevención de la utilización del sistema financiero para el blanqueo de capitales o la financiación del terrorismo, y por la que se modifican las Directivas 2009/138/CE y 2013/36/UE, que obliga a los agentes privados tales como bancos, auditores y notarios a

[454] Véase Ambos, 2020.

informar sobre aquellas transacciones sospechosas, así como a establecer procedimientos de evaluación de riesgos (razón por la cual, especialmente las entidades bancarias cuentan con sus propios sistemas de detección de riesgos que hacen saltar alarmas en casos de sospecha).

No obstante, procede hacer especial mención a la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, que regula de forma específica este tipo de herramientas (a la que posteriormente se hará especial mención al hacer referencia al uso de los sistemas de evaluación de riesgos en el ámbito judicial); así como a la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, que califica este tipo de sistemas como de alto riesgo en su artículo 6, en relación con su Anexo III (punto 6.e) y, por ende, les resulta de aplicación toda la normativa ya mencionada al respecto que, no obstante, todavía no está vigente.

Además, se han implementado sistemas de detección de riesgos tales como el "*Visa Information System*" (VIS), cuya gestión operativa corre a cargo de la Agencia de la UE para Sistemas Informáticos a Gran Escala (eu-LISA), que permite a los Estados Schengen intercambiar datos sobre las solicitudes de visados. Dicha herramienta cuenta con un sistema informático central y una infraestructura de comunicaciones que conecta dicho sistema con los sistemas nacionales y los vincula con los consulados existentes en países no pertenecientes a la UE y con todos los pasos fronterizos exteriores de los Estados Schengen. Así, siendo uno de los propósitos del sistema la mejora de la seguridad (en concreto, la prevención, detección e investigación de delitos terroristas y otros delitos graves), los datos de las solicitudes de visados de corta y larga estancia se analizan, se cruzan y se procesan con un conjunto de "indicadores de riesgo", que combinan información relativa al rango de edad, sexo, nacionalidad, país y ciudad de residencia, el propósito del viaje, el Estado Miembro de entrada a la UE y la ocupación, entre otros,

pudiendo incluso realizar una comparación biométrica, principalmente de huellas dactilares, con fines de identificación y verificación.[455]

Asimismo, el Sistema Europeo de Información y Autorización de Viajes (ETIAS), es una herramienta (programada para entrar en completo funcionamiento en el año 2022), tiene como finalidad someter, mediante el cruce de datos, a un control de seguridad exhaustivo a cada solicitante de entrada en el espacio Schengen que, si bien no tenga necesidad de visado, provenga de un país externo a la UE[456], para determinar si se le permite el acceso o no, con el objetivo de asegurar que no resulta una amenaza para la seguridad, principalmente (y la salud pública, también, por el riesgo epidemiológico).

Y especial mención merece el programa SyRI ("*System Risk Indicator*"), un sistema de detección de riesgo de comisión de fraude frente a la Seguridad Social y la Hacienda Pública empleado por el Gobierno holandés, respecto del que un Tribunal de Distrito de La Haya (Países Bajos)[457] falló que el derecho a la privacidad debía prevalecer sobre el interés público, siendo que el uso del mismo resultaba desproporcionado en relación al fin que se quería conseguir con él.

Por su parte, en el ámbito externo a la UE, muy interesante y necesario es hacer referencia a la herramienta del gobierno chino de creación de perfiles denominada "sistema de crédito social de China" ("*China's social credit system*"), la más intrusiva y completa que se está proyectando en el mundo (al menos, que se sepa) y que, desde luego, tiene todos los visos de conculcar frontalmente los principios, valores y derechos sobre los que se construye no solo el marco jurídico europeo sino también el de la mayoría de Estados democráticos y de Derecho.

Y es que ya en el año 2014 (con un objetivo de implementación a lo largo del año 2020) el gobierno chino reveló[458] que estaba desarrollando un sistema de IA, de uso obligatorio,

[455] Véase Comisión Europea, 2008.
[456] Hay 62 países que no forman parte de la UE cuyos nacionales pueden entrar en el espacio Schengen sin necesidad de visado.
[457] Véase Alston, 2019.
[458] Véase China Copyright and Media, 2014.

para controlar el comportamiento de sus ciudadanos y empresas, en todos los ámbitos de la vida, y poder así analizar y calificar o puntuar la fiabilidad de los mismos.

En la actualidad, en China no existe un único sistema de crédito social, siendo que los gobiernos locales tienen los suyos propios (como por ejemplo, el de la ciudad de Suzhou[459]), más limitados, existiendo asimismo versiones privadas no oficiales que operan empresas como Zhima Credit de Ant Financial (empresa de pagos surgida de Alibaba), más conocida como Sesame Credit. Así, el objetivo del gobierno central chino no es otro que el de lanzar un sistema de crédito social único a nivel nacional que dote a las empresas de un código de crédito social unificado, y a los ciudadanos de un número de identidad, todos vinculados a un registro permanente, para facilitar el flujo de información.

En virtud de ello, el gobierno chino pretende recopilar millones de datos e informaciones de sus ciudadanos a través de múltiples fuentes y luego analizarlos mediante algoritmos absolutamente opacos (al menos por el momento), logrando así fijar la calificación *rating* de cada uno de ellos. Los factores a tener en cuenta por el sistema para establecer dicha puntuación o crédito social, aparentemente, son múltiples y variados, y entre otros: impagos de multas, malos comportamientos en el transporte público, impagos de servicios de suministros, comisión de infracciones administrativas (tales como pasar un semáforo en rojo, por ejemplo), peleas vecinales, etc.

Algunos informes, además, hacen referencia a la posible existencia de una lista negra que haría perder ciertos derechos a los ciudadanos incluidos en ella (tales como por ejemplo, la posibilidad de reservar un billete de tren o avión, comprar propiedades o solicitar un préstamo). No obstante, también se apunta a que una calificación positiva pueda implicar descuentos y beneficios (tales como, por ejemplo, un proceso burocrático simplificado). Y todo ello basado en la "aparente" necesidad de generar confianza, bajo la creencia de que "*mantener la confianza no se recompensa lo suficiente, y sin embargo los costes de romper la confianza tienden a ser bajos*", tal y como se dispone en el documento emitido por el gobierno chino en 2014 al que se ha hecho referencia con anterioridad.

[459] Véase Chiu, 2020.

La utilización por parte del gobierno, con carácter obligatorio, de un sistema como el expuesto, sin duda abre la puerta a vulneraciones masivas de los derechos de sus ciudadanos, especialmente en caso de que los datos (que pueden resultar erróneos en muchas ocasiones y abocar a situaciones muy injustas para los ciudadanos) y los algoritmos empleados permanezcan secretos y, sobre todo, en caso de que los ciudadanos no tengan capacidad real de denunciar las conculcaciones de sus derechos ante ninguna instancia judicial o gubernamental. No obstante, siendo que China se ha mostrado aparentemente concienciada con los posibles efectos dañinos que puede entrañar el uso de sistemas de IA que no cumplan con unos mínimos éticos, tal y como se ha anunciado en puntos anteriores, hace pensar que dicho país merece el beneficio de la duda y que habrá que esperar a que el mencionado sistema de crédito social lleve un tiempo en funcionamiento para poder determinar y analizar sus efectos reales.

Y es que tal y como se puede observar, ingentes cantidades de datos, a través de sistemas de IA, pueden ser cruzados con los datos provenientes de otras fuentes y posteriormente analizados y procesados para llevar a cabo la elaboración de perfiles. Así, existen herramientas dotadas de tal tecnología capaces de buscar información de manera eficaz entre innumerables cantidades de datos heterogéneos, provenientes de múltiples fuentes y formatos, siendo Gotham una de ellas, creada por la empresa de *software* estadounidense Palantir y empleado por EUROPOL desde el año 2016[460] y por diversas fuerzas policiales de toda Europa que han firmado contratos millonarios con tal compañía para la compra de herramientas de análisis masivos de datos (entre otras, la Dirección General de Seguridad Interior de Francia -DGSI-, que en 2016 contrató los servicios de Palantir por 10 millones de euros para luchar contra el terrorismo[461], aunque en 2018 optó por los servicios de "*Cluster Data Intelligence*", una alternativa 100% francesa).[462] Por su parte, la policía de Hessen (Alemania) emplea una versión adaptada de la herramienta "*Gotham*" de Palantir denominada "*Hessendata*", capaz de detectar patrones y crear perfiles y gráficos sobre posibles sospechosos de terrorismo, utilizando siete fuentes de datos, incluidas bases de datos policiales y perfiles de Facebook.[463] Y, por su parte, en Estados Unidos la

[460] Véase Parlamento Europeo, 2020.
[461] Véase Cheminat, 2016.
[462] Véase Boulestin, 2018.
[463] Véase Herberg & Lindhoff, 2020.

herramienta de creación de perfiles Operation LASER clasifica a las personas con base en sus circunstancias, vida personal e interaccion con el sistema judicial para detectar perfiles que requiren mayor vigilancia.[464]

Dichos programas de *software*, como ya se ha advertido, bucean en numerosas fuentes, tales como las redes sociales, que son un oasis de información, ya que sus millones de usuarios discuten públicamente sobre emociones, eventos e innumerables temáticas, con contenido de distribución a tiempo real, y los mensajes a menudo muestran coordenadas geotemporales precisas. Asimismo, los antedichos sistemas extraen, analizan y procesan información de los registros económicos, los registros de inteligencia y policiales, y otros datos de fuente abierta, que pueden incluir información de Internet y de espacios públicos, etc. A pesar de ello, existen millones de datos que pueden permanecer ocultos incluso para los programas de *software* más sofisticados, habida cuenta de que los delincuentes ya se preocupan de que estos se hallen encriptados, tal y como ocurre con las comunicaciones end-to-end (E2E) y las establecidas a través de la denominada *Dark Web*, así como las transacciones realizadas con criptomonedas o a través de sistemas de *Blockchain*, pero eso ya es otra problemática que sería merecedora de una tesis doctoral independiente.

No obstante, una cuestión a tener en cuenta, en relación con este tipo de sistemas, tal y como se pone de relive desde EUROPOL[465], es que actualmente debe fijarse el foco en el uso de la tecnología 5G, que permite conexiones de información y datos de los dispositivos conectados significativamente más rápidas, lo que sin duda puede ser aplicado a los *software* antedichos y aportar grandes beneficios. No obstante, tal nueva tecnología también crea nuevos desafíos para las fuerzas de seguridad, entre otros, la dificultad para identificar y localizar a los usuarios, ya que la puesta en marcha de las redes 5G implica una fragmentación de la información, por lo que es posible que esta no esté disponible o no sea tan fácilmente accesible para los cuerpos policiales, resultando más necesaria que nunca la cooperación entre los distintos proveedores de redes tanto nacionales como internacionales con fines de seguridad. Además, a través de la tecnología 5G los dispositivos pueden comunicarse directamente entre sí sin tener que usar la red central del

[464] Véase Uchida & Swatt, 2013.
[465] Europol, 2019.

operador, lo que complica muchísimo la posibilidad de recuperar los datos de la comunicación.

Como es sabido, los datos recogidos a través de tales sistemas, que son muchísimos, son procesados por herramientas de IA que se encargan de generar las oportunas alertas a los cuerpos de inteligencia, defensa o policía, para que puedan así actuar en consecuencia.

Sin embargo, y a pesar de las bonanzas expuestas, al igual que ocurría en el caso de los sistemas de mapeo predictivo, procede poner de manifiesto que estas herramientas de IA se enfrentan a desafíos éticos y legales de enormes dimensiones. Y es que, aunque la mayoría de ellas se consideran legítimas por tender a garantizar la seguridad pública, el eterno debate "seguridad *vs* resto de derechos fundamentales" cobra especial relevancia en este ámbito.

A modo de ejemplo, a finales de 2013, en Chicago, cuyo Departamento de Policía hacía uso del sistema CompStat, Robert McDaniel, un varón de raza negra de 22 años que residía al sur de la ciudad (en un barrio habitado, en gran medida, por personas de raza afroamericana), recibió por sorpresa la visita de un agente policía para advertirle de que no cometiera delitos, si bien este ni tenía antecedentes penales por delitos violentos ni había tenido contacto reciente con las fuerzas policiales. La explicación de la visita, no obstante, fue que el Sr.McDaniel era una de las aproximadamente cuatrocientas personas incluidas en la "lista caliente" de potenciales delincuentes del Departamento de Policía de Chicago. No obstante, la transparencia del sistema brilló por su ausencia, y por ende, la justificación de la visita, quedó en el aire. [466]

Y es que, si bien está proliferando el uso de herramientas de predicción identificativa, gran parte de estas son difíciles de evaluar o auditar puesto que son opacas y no cumplen con los estándares de transparencia y rendición de cuentas, por ejemplo, que exige la guía ética de IA publicada por la UE y la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la

[466] Véase Gorner, 2013.

Unión, resultando casi imposible llevar a cabo el examen de su legalidad y del cumplimiento de los requisitos de necesidad y proporcionalidad que se desprenden de la regulación europea y española.

Así, surgen varias preguntas que requerirían una respuesta clara anterior a la puesta en marcha de los sistemas analizados: ¿en el ámbito de la UE y de España, la injerencia en el derecho a la intimidad y a la protección de datos de una persona resulta justificada por la mera existencia de un peligro potencial y no concreto? ¿pueden calificarse las herramientas de identificación predictiva como técnicas de investigación prospectiva? ¿quién se encarga de seleccionar, procesar y evaluar los datos? ¿cómo se llevan a cabo tales operaciones? ¿qué datos y qué categorías de los mismos se están seleccionando, procesando y analizando? ¿cómo se controla que tales operaciones no perpetúen sesgos y/o resulten discriminatorias? ¿debe incluirse en tales procesos a las autoridades supervisoras de protección de datos y a los órganos de control democrático?, entre otras.

En relación con ello, la Agencia de Derechos Fundamentales de la UE ha recomendado posponer el uso de las técnicas de elaboración de perfiles hasta que exista una regulación específica, clara y determinante que permita que estas sean empleadas con todas las garantías para los derechos fundamentales de los ciudadanos[467], y no puedo estar más de acuerdo con tal recomendación.

En concreto, la mencionada Agencia ha dispuesto:

"*-la elaboración de perfiles implica categorizar a las personas según sus características;*

-para recopilar y procesar datos personales, las autoridades policiales y de gestión de fronteras deben asegurarse de que la recopilación y el procesamiento de tales datos tengan una base legal, un objetivo válido y legítimo y sean necesarios y proporcionados;

-las características personales protegidas tales como la raza, el origen étnico, el género o la religión pueden ser factores que las autoridades encargadas de hacer cumplir la ley y

[467] Véase Agencia de Derechos Fundamentales de la UE, 2018.

los guardias fronterizos tengan en cuenta para ejercer sus funciones, pero no pueden ser la única o principal razón para señalar a un individuo;

-la elaboración de perfiles que se base única o principalmente en una o más características personales protegidas equivale a discriminación directa y, por lo tanto, vulnera los derechos y libertades del individuo y es ilegal."[468]

Y añade:

"*-Para ser legales, las detenciones y remisiones a controles fronterizos de segunda línea deben basarse en motivos de sospecha razonable y objetiva;*

-las características personales protegidas pueden utilizarse legítimamente para la elaboración de perfiles. Sin embargo, para evitar la discriminación, también debe haber motivos fundados y razonables de sospecha basados en información distinta a las mencionadas características protegidas."[469] (…)

En relación con ello, Irakli Beridze, Director del UNICRI Centre for Artificial Intelligence and Robotics en una entrevista concedida al medio Geospatial World en septiembre del 2019, respecto de la imprescindible necesidad de creación de un marco legal específico para la vigilancia policial predictiva, en especial en términos de intercambio de datos y acceso a los mismos, aseguró: "*Creo que a medida que la IA se vuelva más efectiva y sofisticada, será empleada para todas las formas de vigilancia, y ello ocurrirá en un futuro no muy lejano. Por lo tanto, es muy importante que los organismos encargados de hacer cumplir la ley se guíen por nociones como justicia, responsabilidad, transparencia y explicabilidad. Y es que si el uso de IA por parte de las fuerzas del orden no es legal y confiable, se ponen en peligro nuestros derechos humanos y se socavan en gran medida los principios fundamentales del Derecho, tales como la presunción de inocencia, la facultad de no autoinculparse y la prueba más allá de toda duda razonable. Para evitar esto, se requerirá algún marco para la aplicación de la ley.*"[470]

[468] Agencia de Derechos Fundamentales de la UE, 2018, pág. 12.
[469] Agencia de Derechos Fundamentales de la UE, 2018, pág. 13.
[470] Hisham, 2019.

Y es que los derechos fundamentales y las libertades individuales en juego son muchos y muy sensibles, y si bien las herramientas analizadas pueden aportar grandes beneficios a la sociedad, su puesta en circulación de forma precipitada, dispersa y poco transparente no ha hecho más que demonizarlas, ya que de entrada su uso ha sido percibido como una amenaza para los ciudadanos más que como una ayuda y ha ocasionado una evidente pérdida de confianza en las autoridades policiales.

En relación con ello, si bien debe tenerse en cuenta, desde luego, que el Derecho no puede regular de forma matemática el peso de cada derecho fundamental en juego en cada caso, sí puede fijar unas pautas y unos criterios claros, en función de lo que decida el poder legislativo en cada momento, ya que, como ha ocurrido en el pasado, la realidad social está yendo muchísimo más rápido que la jurídica y eso resulta altamente peligroso, puesto que pone en jaque los valores y derechos en los que se ampara la UE. Y ello es lo que pretende conseguirse con la normativa ya existente, a saber, la LO 7/21, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales; y con la futura, en virtud de lo dispuesto en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión.

a.3) Sistemas de evaluación de riesgos individuales (“*Decision making or Risk assessment relating to individuals”)*

Los sistemas de evaluación de riesgos individuales son aquellas herramientas que, a través de la IA, mediante el análisis de distintas variables, pueden identificar el nivel de probabilidad de que un individuo concreto lleve a cabo un determinado comportamiento futuro, siendo su principal objetivo no solo el de pronosticar tales eventos venideros sino también el de gestionar el peligro que puedan implicar.

En el ámbito policial, merece especial atención la herramienta de *Machine Learning* HART, desarrollada a partir de una colaboración entre la Universidad de Cambridge (Reino Unido) y la policía de Durham Constabulary (Reino Unido), y empleada por esta última

con el fin de asistir a sus agentes a la hora de decidir si un sospechoso detenido debe quedar en libertad, permanecer en custodia o ser seleccionado para participar en un programa de rehabilitación local (el denominado "*Constabulary's Checkpoint programme*"), y todo ello mediante el análisis del riesgo de que este delinca de nuevo en el plazo de los dos años posteriores si resulta liberado.

Es interesante saber que el antedicho sistema se nutre de un historial de aproximadamente cien mil casos ocurridos durante un período de cinco años (2008-2012) y tiene en consideración hasta treinta y cuatro factores distintos para elaborar su pronóstico, la mayoría de los cuales hacen referencia al historial delictivo del sujeto analizado.[471]

Así, HART, a través del análisis de los múltiples datos con los que cuenta, clasifica a los sospechosos detenidos por la policía según su nivel de riesgo: alto (si se prevé que cometan un nuevo delito grave en los próximos dos años), moderado (si se prevé que cometan un nuevo delito menos grave en los próximos dos años, lo cual implica la posibilidad de ser elegidos para entrar en el mencionado programa de rehabilitación "*Constabulary's Checkpoint programme*") y bajo (si se prevé que estos no cometan nuevos delitos durante los dos años siguientes).

No obstante, este tipo de sistemas, empleados por los cuerpos policiales de todo el mundo, ha sido duramente criticado por su extendida opacidad y, especialmente, por su posible trasfondo discriminatorio.

En concreto, a modo de ejemplo, el analizado sistema HART fue objeto de investigación por la organización *pro* derechos humanos Big Brother Watch (BBW)[472], que denunció que, por un lado, los datos de los que se nutría eran suplementados por los de una base de datos de la agencia Experian[473] denominada "*Mosaic*", obtenidos a partir de la creación de perfiles de cincuenta millones de adultos de Reino Unido y, por otro lado, llevaba a cabo el análisis de riesgos sesgado y discriminatorio. Y es que, entre los treinta y cuatro factores que empleaba la mencionada herramienta de IA para realizar su pronóstico, se hallaba

[471] Véase Oswald, Grace, Urwin, & Barnes, 2018, págs. 223-250.
[472] Véase Big Brother Watch, 2018.
[473] Una agencia que se dedica a recopilar y vender información sobre consumidores.

información relativa al código postal de las personas analizadas, lo que sin duda podía hacer aumentar los prejuicios hacia aquellos que vivían en determinadas zonas por el mero hecho de residir allí. No obstante, como consecuencia de ello, y con el fin de mejorar la calidad del contenido y demostrar su compromiso con los derechos de los ciudadanos, la policía de Durham Constabulary decidió eliminar del sistema el campo de código postal principal (que incluía los cuatro primeros dígitos de los códigos postales).[474]

A modo de ejemplo, también, en Filadelfia (EEUU), Darnell Gates salió de prisión en 2018 tras haber cumplido condena por haber introducido un automóvil en una vivienda en 2013 y haber amenazado violentamente a su expareja. No obstante, al quedar en libertad, las autoridades le requirieron para que acudiera una vez a la semana (posteriormente cada quince días y, finalmente, cada mes) a una "oficina de libertad condicional" sin especificarle el porqué, aunque la verdadera razón era que había sido calificado como de "alto riesgo" por un algoritmo, circunstancia de la que fue por primera vez informado por unos periodistas del periódico The New York Times que decidieron entrevistarlo[475], lo cual resulta absolutamente inaceptable.

Sin embargo, tales sistemas de IA también están aportando grandes avances y utilidades. Así, en agosto de 2016, el Condado de Allegheny (Pensilvania, EEUU) se convirtió en la primera jurisdicción de Estados Unidos en permitir que un algoritmo de análisis predictivo ofreciera evaluaciones de riesgo relativas a posibles situaciones de vulnerabilidad de menores en el ámbito familiar, con el fin de ayudar a los agentes de policía a identificar a aquellas familias concretas que necesitaban más intervención[476], lo cual recibió muy buenas críticas. Tal herramienta de IA, desarrollada por Vaithianathan y Putnam-Hornstein, es propiedad del mencionado Condado -que la somete a constante revisión-, su funcionamiento es público, sus criterios -seleccionados por los funcionarios locales- se describen en publicaciones académicas, y ha sido evaluado por expertos independientes, lo que ha provocado que la mayoría de defensores de padres y niños y de los derechos civiles hayan aplaudido su incorporación en el día a día policial. En relación con ello, la

[474] Véase Burgess, 2018.
[475] Metz & Satariano, 2020.
[476] Véase Hurley, 2018.

organización The American Civil Liberties Union (ACLU) de Pennsylvania[477] manifestó al periódico The New York Times: "*Creo que están poniendo controles importantes en el proceso.* (…) *Lo están usando solo para inspectores, para decidir qué llamadas investigar, no para retirar a un niño de su familia.*"[478], y Brett Drake, profesor de la "*Brown School of Social Work*" de la Universidad de Washington (St. Louis, Missouri, EEUU), aseguró al mismo periódico que "*Dados los primeros resultados de Pittsburgh, el análisis predictivo parece una de las innovaciones más interesantes en protección infantil en los últimos veinte años*."[479]

Así, cada vez que un/a agente de policía de dicho condado recibe una llamada o realiza una entrevista que le hace detectar la posible existencia de peligro familiar para un menor, además de efectuar una evaluación humana del riesgo con base en la experiencia, los conocimientos, los protocolos prefijados etc, puede consultar con la herramienta de IA para obtener así una especie de "segunda opinión". Tal sistema, que se basa en un análisis estadístico de las llamadas de los cuatro años anteriores, utilizando más de cien criterios contenidos hasta en ocho bases de datos distintas (prisiones, servicios psiquiátricos, servicios sociales y centros de tratamiento de drogas y alcohol, entre otros), tarda tan solo unos segundos en mostrar en la pantalla una barra de color vertical que va desde un 1 verde (riesgo más bajo) en la parte inferior hasta un 20 rojo (riesgo más alto) en la parte superior. Así, en caso de que el resultado arrojado por el sistema sea de riesgo alto, el agente de policía encargado del caso, con alta probabilidad decidirá abrir una investigación y enviar a la vivienda del menor a un supervisor para que compruebe, *in situ*, cuál es la situación del mismo.

Asimismo, en Houston (EEUU), investigadores de la University of Texas Health Science Center, financiados por el National Institute of Justice (NIJ) han creado algoritmos para analizar la victimización de los ancianos, especialmente en relación con las estafas, y la idea es que tales algoritmos se puedan emplear por los profesionales para determinar de

[477] Una organización sin ánimo de lucro dedicada a defender y proteger los derechos individuales y libertades personales.
[478] Hurley, 2018.
[479] Hurley, 2018.

manera confiable la probabilidad de que se esté produciendo una estafa (especialmente *on line*) e intervenir lo más rápido posible.[480]

En España, en cumplimiento de lo establecido en la Ley Orgánica 1/2004, de 28 de diciembre, de Medidas de Protección Integral contra la Violencia de Género, la Secretaría de Estado de Seguridad del Ministerio del Interior, el 26 de julio del 2007 puso en funcionamiento la herramienta de IA desarrollada por la Policía Nacional, denominada "Sistema de Seguimiento Integral en los casos de Violencia de Género" (Sistema VioGén), con los objetivos de aglutinar a las diferentes instituciones públicas con competencias en materia de violencia machista; integrar toda la información de interés necesaria; hacer predicción del riesgo de cada individuo implicado en un posible caso de violencia machista (agresores y víctimas) y, atendiendo al nivel del mismo, realizar seguimiento y protección a las víctimas en todo el territorio nacional; y efectuar una labor preventiva, emitiendo avisos, alertas y alarmas, a través del "Subsistema de Notificaciones Automatizadas" cuando se detecte alguna incidencia o acontecimiento que pueda poner en peligro la integridad de la víctima.[481] Con el nuevo protocolo establecido en el año 2019, además, el sistema evalúa el riesgo de los menores víctimas de violencia de género, que es calificado como "bajo", "medio", "alto" o "extremo", de forma independiente al de la madre, con aplicación, en su caso, de las medidas de protección que se consideren convenientes.

Según datos del Ministerio del Interior, desde la puesta en funcionamiento de tal herramienta la reincidencia de las agresiones machistas ha disminuido un 25%[482], lo cual desde luego es una buena noticia y una motivación para seguir trabajando en herramientas como la expuesta. Sin embargo, el hecho de que la calidad de tal sistema no conste auditada y certificada por ningún organismo público especializado y que la transparencia sea incompleta (puesto que no consta qué peso otorga el algoritmo a cada factor de evaluación, por ejemplo), resulta altamente peligroso.

[480] Véase Dyer & Burnett, 2014.
[481] Véase Gobierno de España, 2007.
[482] La Vanguardia, 2015.

No obstante, y siendo que las herramientas descritas en el presente apartado y, especialmente, sus posibles beneficios e implicaciones jurídicas serán analizados con más profundidad al hablar de su uso en el ámbito judicial, no procede extender más su examen.

3.2.1.3. Ámbito judicial. Sistemas de justicia predictiva.

Tal y como se ha venido exponiendo, las herramientas de evaluación de riesgos pueden ser definidas como aquellos sistemas que, a través de la IA, identifican potenciales peligros y analizan la probabilidad de que estos se manifiesten, estando diseñadas, principalmente, con el objetivo de informar y asistir (no sustituir, al menos por el momento) a los seres humanos -como tales, limitados- en la toma de decisiones.[483]

En relación con ello, es interesante poner de manifiesto que, durante los últimos años, tales sistemas de IA han sido introducidos en diversos sistemas judiciales de todo el mundo con la finalidad de predecir ciertos comportamientos humanos futuros y ayudar así a los tribunales en la toma de decisiones.

En el ámbito penal, las razones que han motivado la implementación de tales sistemas son principalmente dos: por un lado, la idea de que su utilización puede ayudar a reducir los sesgos humanos, al basarse (aparentemente) en datos objetivos; y, por otro lado, la idea de que su uso promoverá una reforma del sistema penal y, además, reducirá la carga de trabajo y los costes de la Administración de Justicia. No obstante, tales objetivos, en un principio muy claros y alentadores, tal y como se verá más adelante, no están dando los resultados esperados.

Es de utilidad saber que, en el procedimiento penal, la evaluación de riesgos suele llevarse a cabo mediante el denominado "modelo actualarial de evaluación de riesgos", que consiste en asignar valores numéricos a cada factor de riesgo tenido en cuenta por el sistema y,

[483] No obstante, es importante remarcar que el término "*RATs*" no siempre hace referencia a herramientas de evaluación de riesgo que emplean IA para cumplir con sus objetivos, sino que también puede referirse a sistemas más simples, que no se nutren de grandes cantidades de datos y que pueden operar a través de un formulario de evaluación utilizado por el departamento de servicios previos al juicio, por una tarjeta de puntuación utilizada por un juez, etc.

posteriormente, hacer una combinación y ponderación de todos ellos a través de un algoritmo para poder así proporcionar puntuaciones o niveles de riesgo.[484]

Los denominados "factores de riesgo" que se tienen en cuenta por los mencionados sistemas para determinar las probabilidades de que tenga lugar cierto acontecimiento futuro, suelen ser, en concreto, características relativas a la persona analizada, tales como, entre otras, los antecedentes penales, los antecedentes policiales y judiciales (especialmente la existencia de órdenes de busca y captura, etc), la edad, el sexo, los cargos penales pendientes, la adicción a sustancias tóxicas o el arraigo (familiar, laboral, etc). Y tales circunstancias, ponderadas por un algoritmo, suelen arrojar como resultado un nivel de riesgo determinado de que algo ocurra: bajo, medio o alto.

En relación con lo expuesto, considero necesario e interesante hacer una previa y breve enumeración de los principales peligros que puede entrañar el uso de los sistemas analizados por parte de la Administración de Justicia que, no obstante, irán siendo analizados de forma más profunda a lo largo de la presente exposición.

En primer lugar, sin lugar a duda, el principal peligro de los sistemas de IA de evaluación de riesgos a los que se está haciendo referencia es la calidad de los datos que emplean. Y es que la existencia de datos de "mala calidad", a los que ya se ha hecho mención con anterioridad en el presente trabajo, puede causar efectos nefastos en cualquier ámbito de aplicación, si bien su incidencia puede resultar especialmente perniciosa en el ámbito judicial.

Así, por un lado, el problema que subyace en muchos de los aludidos sistemas de IA es que estos están basados en información generada durante periodos históricos en que se llevaban a cabo procedimientos policiales "dudosos" (la denominada "*dirty police*") que, en muchos casos, empleaban datos incorrectos, obtenidos incluso de modo ilegal o con prejuicios sociales, raciales o de otra índole, por lo que si tales datos informan y nutren los sistemas actuales de predicción de riesgos, resulta muy difícil o incluso imposible evitar que los

[484] Véase Desmarais & Lowder, 2019, págs. 4-5.

efectos perjudiciales de aquellos antiguos métodos ilegítimos o sesgados se repitan y se perpetúen.[485]

A modo de ejemplo, si durante una larga época se cometieron detenciones de forma indiscriminada en zonas donde residían principalmente personas de raza negra, los datos que se generaron en su día y que ahora se emplearán por los algoritmos estarán indudablemente sesgados en tal sentido, por lo que las predicciones del sistema resultarán, inevitablemente, discriminatorias, al indicar seguramente que existen más probabilidades de que cometan delitos aquellas personas que residen en zonas donde suelen vivir ciudadanos de color que las que residen en zonas donde habitualmente viven ciudadanos blancos.

No obstante, actualmente empieza a extenderse una fuerte concienciación sobre tal problemática que ya se puso de manifiesto, principalmente, con el estallido del escándalo del programa de estadísticas delictivas conocido como CompStat[486], y se están empezando a adoptar medidas para corregirla.

Por otro lado, resulta evidente que, con anterioridad al uso de la IA, la totalidad de las decisiones judiciales eran tomadas por jueces y, siendo que estos son humanos y que toda decisión tomada por un ser humano es susceptible de contener errores y sesgos (a pesar de la enorme profesionalidad de la mayoría de los miembros de la carrera judicial), los sistemas judiciales inevitablemente estaban plagados de ellos. No obstante, si bien podría pensarse que la utilización de algoritmos podría traer el fin a tal problemática (habida cuenta, sobre todo, de que ello es técnicamente posible), lo cierto es que en la actualidad el uso de algoritmos en el ámbito judicial, en muchas ocasiones, en vez de servir para mejorar el sistema, no está haciendo más que integrar y perpetuar la totalidad de los errores y sesgos contenidos en todas las decisiones pasadas que son introducidas en los mismos, siendo que su capacidad multiplicadora es altamente peligrosa, por lo que en mi opinión, en múltiples casos se ha creado un efecto contrario al que se pretendía conseguir.

[485] Véase Lum & Isaac, 2016.
[486] Véase Eterno & Silverman, 2012.

Y es que los sesgos sistemáticos, la omisión o censura de datos o la mala gestión o concepción de su relevancia por parte de los *software* empleados, pueden llevar a los cuerpos policiales y judiciales a tomar decisiones incorrectas o, incluso, discriminatorias, tal y como se advirtió en el informe publicado en 2019 por el AI Now Institute de la Universidad de Nueva York (EEUU)[487].

En segundo lugar, tal y como se irá viendo a lo largo del presente punto, uno de los mayores problemas que entrañan los sistemas de IA de evaluación de riesgos empleados en el ámbito judicial, al igual que ocurre con la mayoría de herramientas que emplean tal tecnología en la actualidad, es la falta de transparencia. Y es que procede poner de manifiesto que el valor de la transparencia en el ámbito analizado es particularmente relevante, habida cuenta de que los propios fiscales y jueces deben entender la información que están recibiendo para poder así, por un lado, tomar las decisiones más correctas y, por otro lado, determinar si el uso de estas resulta apropiado y responde o no a los estándares jurídicos legalmente establecidos. Y, asimismo, los ciudadanos tienen derecho a conocer el porqué, en su caso, de la toma de decisiones que les afectan y están basadas en los resultados de dichos sistemas.

Así, a modo de ejemplo, en San José (California, EEUU), donde se emplea uno de los mencionados sistemas de IA en los denominados "*arraignment hearings*" (audiencias de acusación), una organización llamada Silicon Valley De-Bug se encarga de entrevistar a la familia de cada acusado, aportar al tribunal la información extraída y compartirla con los abogados de la defensa como una especie de contrapeso a los algoritmos[488], de forma preventiva, puesto que en caso contrario, estos pueden tener una elevada influencia en la toma de decisiones del tribunal sin apenas posibilidad de contradicción.

Y es que, en conclusión, uno de los riesgos principales y potencialmente más nocivos para los ciudadanos asociado a este tipo de sistemas de IA es el de la posible vulneración de derechos fundamentales mediante su empleo, siendo especialmente sensibles la libertad (artículo 6 de la Carta de Derechos Fundamentales de la UE y artículo 17 de la Constitución Española), la privacidad y la protección de datos personales (artículos 7 y 8 de la Carta y

[487] Richardson, Schultz & Crawford, 2018.
[488] Véase Jayadev, 2019.

artículo 18 de la Constitución), la igualdad y la no discriminación (artículos 20 a 23 de la Carta y artículo 14 de la Constitución) y, especialmente, la presunción de inocencia (artículo 48 de la Carta y artículo 24 de la Constitución)[489].

Dicho lo expuesto, es interesante poner de manifiesto que, si bien la utilización de herramientas de evaluación de riesgos en el ámbito judicial todavía está muy poco extendida en la UE, y por ende, en España, estas están especialmente presentes en EEUU, donde su uso se ha multiplicado en los últimos años[490], aunque ya llevan más de una década en funcionamiento.

En España, no obstante, especialmente interesante es hacer referencia al sistema de IA denominado RisCanvi, empleado en las prisiones de Cataluña (y, por ende, en los tribunales) desde el año 2009, que tiene como finalidad principal medir el riesgo de reincidencia de los presos para ayudar a las autoridades a decidir si estos están listos o no para salir de prisión (a pesar de que también se emplea para medir el riesgo de quebrantamiento de condena, de violencia intrainstitucional -contra otros presos o contra funcionarios de prisiones-, y de violencia auto-dirigida -autolesiones, suicidio, etc-).

Así, cuando un interno solicita determinados permisos (entre otros, la libertad condicional), el juez que debe decidir al respecto recibe un informe del centro penitenciario que contiene una serie de informaciones y elementos sobre los que fundamentar su decisión que, en parte, es elaborado por un algoritmo.

Y es que la mencionada herramienta analiza y valora cuarenta y tres factores distintos para arrojar un resultado de riesgo bajo, medio o alto de reincidencia (con especial mención a la reincidencia violenta), entre los que se hallan la edad del interno en el momento de cometer el delito, su género, su origen, el delito base violento, la eventual intoxicación durante la realización del mismo, la duración de la pena, los conflictos con otros internos, los eventuales problemas relacionados con la ocupación, la existencia de expedientes

[489] Y, de forma indirecta, el derecho a la dignidad de la persona, reconocido en el artículo 1 de la Carta y en el artículo 10 de la Constitución.
[490] En 2019, un análisis de noventa y una jurisdicciones de EEUU detectó que más de dos tercios utilizaban una evaluación de riesgos previa al juicio. Véase Pretrial Justice Institute, 2019.

disciplinarios, evasiones o fugas, el eventual desajuste infantil, la falta de recursos económicos, fla alta de apoyo familiar y social, regresiones de grado pasadas, ausencia de planes viables de futuro e irresponsabilidad. [491]

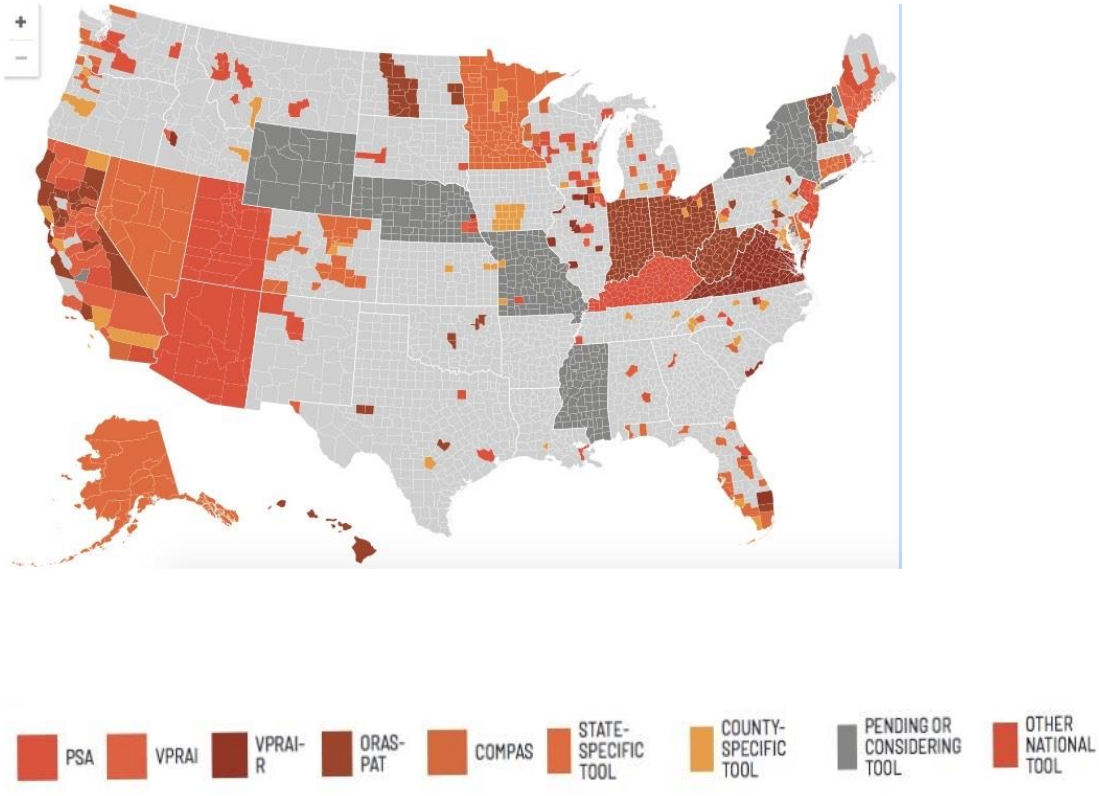
En relación con ello, el Departament de Justicia de la Generalitat de Cataluña defiende que tal sistema es muy valioso, que tiene elevados grados de éxito y que en ningún caso sustituye al juicio humano, sino que simplemente lo orienta. [492] No obstante, tal consideración no es compartida en todos los sectores implicados.

Y es que, si bien la mencionada herramienta se reputa eficaz en la predicción de las tasas de bajo riesgo de reincidencia, lo cierto es que según un informe elaborado en 2015 por el Centro de Estudios Jurídicos de la Generalitat de Cataluña,[493] un 82% de los internos clasificados con riesgo alto o moderado no volvieron a cometer delitos violentos, lo cual es altamente peligroso. Y, en especial, habida cuenta de lo afirmado por Daniel Varona, Catedrático de Derecho Penal y magistrado suplente en la Audiencia Provincial de Gerona que aseguró que: "*RisCanvi se utiliza de forma habitual para denegar permisos*" (…) "nos *encontramos a menudo que, aunque la junta de tratamiento de prisión defienda el permiso, si hay un riesgo alto o medio en Riscanvi el fiscal lo utiliza para recurrir y el juez de vigilancia le da la razón*".[494]

Y lo peor, bajo mi punto de vista, no es que el sistema falle (que, por supuesto, también), sino que su calidad no esté ni auditada ni certificada por ningún organismo público especializado, la transparencia brille por su ausencia y, aun así, se emplee por las autoridades como fundamento de sus decisiones. Y es que no hay forma de conocer qué peso tiene cada uno de los cuarenta y tres factores valorados por el sistema para arrojar sus resultados[495], lo cual, tal y como desarrollaré más adelante, bajo mi punto de vista supone una clara vulneración del derecho a la defensa y puede entrañar la de otros derechos fundamentales.

[491] Ferez-Mangas & Andrés-Pueyo, 2018, pág. 6.
[492] Saura & Aragó, 2021.
[493] Véase Generalitat de Catalunya, 2015.
[494] Saura & Aragó, 2021.
[495] Saura & Aragó, 2021.

En EEUU, por su parte, las organizaciones Media Mobilizing Project de Filadelfia (Pensilvania, EEUU) y MediaJustice de Oakland (California, EEUU) recientemente elaboraron una base de datos nacional que muestra en qué jurisdicciones de EEUU se emplean herramientas de IA de evaluación de riesgos y, en su caso, cuáles, tal y como se muestra en el siguiente mapa:



496

Al respecto, es interesante poner de relieve que en dicho país norteamericano, tales herramientas de evaluación de riesgos tienen un especial uso e impacto en la fase previa al juicio penal ("*Pretrial Risk Assessment Tools*"), donde son empleadas para asistir en la toma de decisiones sobre la situación personal y/o, en su caso, tratamiento de los presuntos delincuentes. No obstante, dichos sistemas de IA se emplean cada vez más, asimismo, en la fase del juicio, a la hora de dictar sentencia, y con posterioridad al juicio, para tomar decisiones relativas a la libertad condicional.

[496] Mapping Pretrial Injustice, 2019.

La razón por la que el uso de tales sistemas ha proliferado tanto en EEUU no es otra que la realidad de sus clamorosos procedimientos judiciales penales, que provocan la saturación de las prisiones de gran parte del país (a finales de 2016, aproximadamente dos millones de adultos se hallaban presos, y otros aproximadamente cuatro millones se encontraban en otras instalaciones correccionales, lo que implica que uno de cada treinta y ocho estadounidenses adultos estaba bajo algún tipo de supervisión pública[497]) y hacen que muchas personas que podrían tener éxito en la comunidad si quedaran en libertad con carácter previo al juicio, permanezcan en prisión debido a su imposibilidad de depositar las fianzas impuestas. Así, las distintas jurisdicciones de EEUU llevan años explorando alternativas a la fianza dineraria, y cada vez ha ido cobrando más fuerza la posibilidad de sustitución de tal medida por la presentación de un informe que deje constancia de la probabilidad que tiene una persona investigada de comparecer ante el tribunal sin una nueva detención antes de que se celebre el acto del juicio.[498]

En relación con ello y, en aras de regular el uso de tales herramientas de IA, algunas jurisdicciones estadounidenses han aprobado leyes al respecto y otras han fijado, a través de sus Tribunales Supremos, pautas de aplicación, si bien todavía existen muchos estados o condados que no tienen ninguna base legal que dé cobertura a la utilización de dichos sistemas, lo cual, en mi opinión, resulta altamente peligroso.

En relación con ello debe hacerse especial mención a la herramienta de evaluación de riesgos empleada, por excelencia, en el ámbito judicial penal estadounidense: COMPAS.

COMPAS es un sistema de IA que fue desarrollado ya en 1998 por una empresa privada, Northpointe Inc., que contiene un algoritmo que, por un lado, calcula la probabilidad de que un acusado reincida y, por otro lado, sugiere qué tipo de supervisión/tratamiento debería este recibir en prisión, y todo ello a partir de la información recopilada sobre su conducta pasada y la realización de un cuestionario[499]. No obstante, el origen de los datos, la totalidad de los factores que se tienen en consideración y el peso que se otorga a cada

[497] Hao, 2019.
[498] Véase Desmarais & Lowder, 2019, pág. 3.
[499] Véase el cuestionario en ProPublica, s.f..

uno de ellos, permanecen ocultos, por lo que resulta imposible entender qué hay detrás de los resultados que el sistema arroja.

En concreto, la empresa Northpointe Inc. hizo público que "*COMPAS es un paquete de software automatizado de apoyo a la toma de decisiones que integra la evaluación de los riesgos y las necesidades de un individuo con otras cuestiones, tales como fallos de sentencia, tratamiento y administración de casos, y resultados de reincidencia*" [500], remitiendo a su página web, donde, dicho sea de paso, no se revelan más que generalidades y se mencionan otras herramientas de evaluación de riesgos que ofrece la compañía (que el 9 de enero de 2017 se fusionó con las otras dos empresas estadounidenses más punteras en el ámbito de la IA para uso penal, Inc.CourtView Justice Solutions Inc. y Constellation Justice Systems Inc., habiendo dado lugar a la actual compañía Equivant[501]), a saber, LSI-R, ORAS y WRNA TRAILER v5 o v6.

Tal y como publica la empresa Equivant en su guía práctica, COMPAS tiene dos modelos principales de riesgo: Riesgo de Reincidencia General ("*General Recidivism Risk*") y Riesgo de Reincidencia Violenta ("*Violent Recidivism Risk*"), y dispone de escalas que miden tanto el riesgo dinámico (factores criminológicos) como el riesgo estático (factores históricos). Asimismo, existen modelos de riesgo adicionales que incluyen la Pantalla de Riesgo de Reincidencia ("*Recidivism Risk Screen*") y la Escala de Riesgo de Liberación Previa al Juicio-II ("*Pretrial Release Risk Scale II*").

Respecto de ello, se pone de manifiesto que el sistema de evaluación de riesgos COMPAS está diseñado para poder ser configurado por los usuarios, respecto de los elementos de decisión, en función del sistema de justicia penal local y en relación a diferentes poblaciones. Así, por ejemplo, los "*Pre-Trial Services*" de cada estado o condado (en el caso de EEUU) pueden optar por utilizar únicamente la Escala de Riesgo de Liberación Previa al Juicio-II ("*Pretrial Release Risk Scale II*") para hacer recomendaciones al tribunal respecto de liberación de una persona investigada antes del juicio, usar posteriormente las escalas de Riesgo de Reincidencia Violenta ("*Violent Recidivism Risk*") y de Riesgo de Reincidencia General ("*General Recidivism Risk Scales*") para clasificar los casos por

[500] Brennan, Dieterich & Ehret, 2009, págs. 22-23.
[501] Véase StreetInsider, 2017.

riesgo de reincidencia, y optar por realizar la evaluación completa (que proporciona una visión integral de la persona para abordar las necesidades de supervisión y tratamiento para rehabilitación) solo en los casos de las personas de mayor riesgo.

Asimismo, desde la compañía Equivant se insiste en afirmar que el sistema COMPAS se va revisando de forma constante, a medida que los conocimientos técnicos van avanzando, con el fin de ir actualizándolo periódicamente para mantenerse al día. Y, como ejemplo de ello, el PRRS-I ("*Pretrial Release Risk Scale-I*") fue modificado en 2019 por el PRRS-II ("*Pretrial Release Risk Scale-II*"), siendo la principal novedad que este último no incluye la edad de la persona analizada como factor de evaluación del riesgo.[502]

No obstante, los algoritmos utilizados por el sistema COMPAS siguen siendo un misterio amparado por el secreto de empresa (a pesar de que, como veremos, la investigación efectuada por ProPublica arrojó cierta luz sobre la información empleada por estos), ya que tal y como manifestó Jeffrey Harmon, en su momento Director General de Northpointe Inc., al periódico The New York Times: "*La clave de nuestro producto son los algoritmos, y son de nuestra propiedad. Los hemos creado y no los publicamos porque, sin duda, son una pieza fundamental de nuestro negocio. No se trata de mirar los algoritmos. Se trata de observar los resultados*".[503]

Y es que, precisamente, uno de los mayores y más complejos debates que se plantean en torno al uso de tales instrumentos de IA en el proceso judicial hace referencia a la colisión del derecho a la defensa de los acusados y a un proceso con todas las garantías (que implica el derecho a conocer todos los elementos que han llevado al fiscal y al juez a tomar una decisión determinada, para poder así, en su caso, rebatirlos), con el derecho al secreto de empresa. Así, resulta evidente que el derecho a la defensa y a un proceso con todas las garantías requeriría la existencia de un sistema de evaluación de riesgos absolutamente transparente y explicable, lo cual entraría en confrontación, no obstante, con el contenido del derecho al secreto comercial, que faculta a las empresas a guardar silencio sobre sus principales hallazgos (en la mayoría de ocasiones amparados por patentes millonarias) para

[502] Gerchick & al, 2019, pág. 1.
[503] Smith, 2016.

mantener su posicionamiento en el mercado y no compartir su *know how* y los resultados de sus inversiones con la competencia.

Tal cuestión, sobre la que todavía no se ha pronunciado el Tribunal Supremo de EEUU, siendo esperable y deseable que lo haga en un futuro no muy lejano debido al enorme impacto y dificultad jurídica que entraña, sí fue analizada y resuelta por el Tribunal Supremo de Wisconsin (EEUU) en el conocido caso "*State v. Loomis*", que merece ser analizado con detenimiento por haberse convertido en el principal referente sobre el uso en el proceso penal de herramientas que evalúan riesgos a través de la IA y sus implicaciones jurídicas, especialmente en relación al derecho a la defensa y a un proceso con todas las garantías.

A principios del año 2013, Eric Loomis, un ciudadano estadounidense, fue detenido por agentes de policía del Estado de Wisconsin (EEUU) acusado de cinco cargos relacionados con un tiroteo llevado a cabo desde un vehículo en la ciudad de La Crosse (Wisconsin, EEUU). Tal individuo negó en todo momento haber participado en el mencionado tiroteo, si bien admitió haber conducido el automóvil objeto de autos esa misma noche, en un momento posterior a los hechos y, con el ánimo de quedar en libertad u obtener una pena atenuada, se declaró culpable de dos de los cargos menos graves de los que se le acusaba: intentar huir de la policía y conducir un vehículo sin el consentimiento de su propietario.

No obstante, durante la vista que se celebró para preparar la sentencia y decidir sobre su libertad o su ingreso en prisión, por un lado, el tribunal dejó claro a Loomis que el mero hecho de haber reconocido dos de los hechos imputados no implicaba que el resto de cargos que sobre él pesaban no fueran a ser analizados en la sentencia que iba a dictarse; y, por otro lado, un oficial del Wisconsin Department of Corrections aportó al tribunal un informe ("*Pre Sentence Investigation*" -PSI-) que incluía una evaluación de riesgos elaborada por el sistema COMPAS y concluía que el acusado tenía un elevado riesgo de reincidir y cometer actos violentos en el futuro, por lo que representaba un alto riesgo para la sociedad.

Tras ello, el tribunal, en su sentencia, consideró culpable a Loomis no solo de los delitos reconocidos sino también de haber sido quien conducía el vehículo de autos cuando su supuesto colega Michael Vang comenzó un tiroteo que podría haber acabado con la vida

de una o más personas. Para llegar a tal conclusión, el tribunal examinó las características individuales del mencionado acusado, tales como que este creció en un entorno caótico y lleno de obstáculos; que no constaba que hubiera hecho un esfuerzo para superar sus adversas circunstancias y aprender a vivir en comunidad; que tenía un gran y contínuo historial de delitos graves; que contaba con una historia laboral discontínua y, además, necesitaba tratamiento por su adicción a las drogas y su comportamiento pasado como delincuente sexual.

En virtud de ello, el tribunal, que entendió que los delitos cometidos eran muy graves y que, por ende, era necesario proteger a la comunidad y someter a Loomis a tratamiento, puso de manifiesto que el sistema COMPAS concluyó que este era un ciudadano de alto riesgo, por lo que le impuso una pena de seis años de prisión y cinco años más de libertad vigilada y rechazó la libertad condicional solicitada por la defensa, lo cual resultó totalmente sorpresivo e inesperado tanto para el condenado como para su Letrado, habida cuenta del reconocimiento parcial de los hechos que había realizado.

Tras ello, la defensa de Loomis solicitó una nueva audiencia con el tribunal de instancia y argumentó que el uso de la herramienta de evaluación de riesgos COMPAS en la sentencia había vulnerado sus derechos, lo cual fue negado por el tribunal, que básicamente respondió que en caso de no haber empleado tal sistema de IA, el fallo hubiera resultado exactamente el mismo, puesto que la principal base de la condena habían sido otros motivos, todos ellos expuestos en la antedicha resolución.

Como consecuencia de lo anterior, no obstante, el Letrado de Eric Loomis, David Thompson, decidió recurrir la sentencia dictada en primera instancia alegando una frontal vulneración de los derechos de su cliente. El Tribunal Supremo de Wisconsinsin embargo, a pesar de las alegaciones efectuadas, decidió confirmar la resolución recurrida y abrir así la puerta al uso por parte de los tribunales de dicho estado del sistema de evaluación de riesgos COMPAS a la hora de dictar sentencia, habiendo, no obstante, sentado unas determinadas bases para ello. Y es que, en términos generales, dicho tribunal puso de manifiesto que los sistemas de evaluación de riesgos basados en pruebas y evidencias, incluido COMPAS, brindan a los jueces y magistrados información muy relevante y valiosa para llegar a conseguir los dos principales objetivos de la sentencia: la protección

de la comunidad y la rehabilitación del acusado o condenado, por lo que esta debe ser empleada, con ciertas limitaciones, a la hora de dictar resolución, siendo que además aporta resultados más eficaces que la mera intuición humana.

En relación con ello, resulta muy interesante exponer las principales alegaciones efectuadas por el Letrado de la defensa (especialmente las realizadas en contra del uso de la herramienta COMPAS por parte del tribunal), y las respuestas que le fue dando el Tribunal Supremo de Wisconsin en su sentencia, que pueden sistematizarse del modo siguiente.

En primer lugar, la defensa alegó que un acusado tenía el derecho constitucionalmente reconocido al proceso debido y con todas las garantías y, por ende, a ser juzgado con base en una información real y precisa. Respecto de ello, puso de manifiesto que el hecho de que COMPAS fuera un sistema patentado por una empresa privada y amparado por el secreto comercial impedía, por un lado, que pudiera cuestionarse la validez científica de la evaluación de riesgos aplicada (ya que no había acceso completo a toda la información) y, por otro lado, la posibilidad de garantizar que se estuviera sentenciando sobre la base de una información precisa.

En relación con ello, el Tribunal Supremo adujo que el uso de COMPAS no vulneraba los derechos alegados por la defensa habida cuenta de que, por un lado, diversos estados que lo empleaban habían realizado estudios de validación concluyendo que era una herramienta suficientemente precisa (así, se puso de manifiesto que la División de Servicios de Justicia Penal del Estado de Nueva York llevó a cabo un estudio que examinó la efectividad y la precisión predictiva de la escala de reincidencia de una evaluación efectuada por COMPAS y concluyó que esta funcionó de manera efectiva y logró una precisión predictiva exitosa); y, por otro lado, el mencionado Tribunal alegó que el simple hecho de que COMPAS tuviera el potencial de predecir de forma incorrecta el riesgo de un delito futuro por parte de un delincuente individual no podía implicar necesariamente que la información fuera inexacta, ya que además los tribunales son conscientes de que la evaluación de COMPAS no es una garantía de un resultado, sino una mera predicción. Y, finalmente, el Alto Tribunal de Wisconsin alegó que el informe COMPAS se basaba en una lista de veintiuna preguntas y respuestas relativas a factores estáticos y en datos disponibles públicamente sobre el historial criminal del acusado, por lo que Loomis tuvo la oportunidad de verificar

y, en su caso, cuestionar, que la información contenida en el mencionado informe era precisa, puesto que tuvo acceso al mismo de igual forma que lo tuvo el tribunal de instancia.

Asimismo, el Tribunal recordó a la defensa que Loomis sabía perfectamente qué preguntas y respuestas planteaba el sistema COMPAS en su formulario y tuvo, por ende, la oportunidad real de impugnarlas. Respecto de ello, además, el tribunal comparó la toma en consideración por los tribunales de los sistemas de evaluación de riesgos (como COMPAS) con la valoración estos hacen de los informes forenses psiquiátricos sobre peligrosidad futura, ya que en ninguno de los casos el acusado puede conocer el exacto peso que se le da a cada factor para llegar a una determinada conclusión, pero ello no le impide, en ningún caso, cuestionarla.

En segudo lugar, la defensa alegó que un acusado tenía el derecho constitucionalmente reconocido al proceso debido y con todas las garantías y, por ende, a no ser sentenciado por razón de su género.[504]

En relación con ello, el Tribunal Supremo de Wisconsin adujo que en la evaluación realizada por el sistema COMPAS no tuvo en cuenta el género del delincuente, siendo que este disponía de dos escalas de riesgo distintas para hombres y para mujeres, lo cual no significaba que la evaluación considerara el género de forma incorrecta. Y es que, al respecto, se puso de manifiesto que dada la evidencia estadística de que los hombres y las mujeres tienen tasas distintas de reincidencia, el uso de herramientas neutrales al género inflaría o desinflaría artificialmente el riesgo y la mezcla de datos daría lugar a predicciones poco precisas y haría que COMPAS fuera menos fiable, puesto que para arrojar resultados exactos, el sistema debe comparar los perfiles de mujeres con los de otras mujeres y los de hombres con los de otros hombres.

En tercer lugar, la defensa alegó que COMPAS fue originariamente diseñado para ayudar a los servicios de prisiones a asignar recursos e identificar las necesidades individuales de los internos, no para asistir a los jueces a la hora de dictar sentencia. Así, el antedicho Letrado entendía que era muy peligroso emplear una herramienta con un propósito distinto

[504] Si bien el mencionado Letrado no sabía la forma exacta en que COMPAS había tenido en cuenta el género masculino de su cliente, estaba seguro de que ello había sido así y por ello lo puso de manifiesto.

de aquel para el que había sido diseñada, existiendo en este caso, además, riesgo de que los tribunales otorgaran un valor excesivo a las predicciones del sistema y dieran demasiada credibilidad a sus conclusiones.

En relación con ello, el Tribunal Supremo de Wisonsin puso de manifiesto que el propósito original de COMPAS no era, en sí mismo, algo que debiera causar impedimento alguno para que los tribunales tuvieran en cuenta sus resultados a la hora de dictar sentencia, ya que aportaba a los jueces información precisa, válida y confiable para determinar la posible reincidencia y evaluar las necesidades de tratamiento del delincuente (de hecho, más precisa que la evaluación judicial por sí sola). Respecto de ello, el mencionado tribunal adujo que el estado de Wisconsin había consultado con los creadores de COMPAS y concluyó que no había contraindicación alguna para que dicho sistema fuera usado para dictar sentencia.

No obstante, el antedicho tribunal aclaró que los resultados de COMPAS no podían ser determinantes del resultado de la sentencia ni podían servir como único factor para condenar a alguien a pena de prisión, pero sí debían tenerse en cuenta como factores relevantes para mejorar la evaluación y la creación de sentencias individualizadas y apropiadas para cada acusado y tomar decisiones tales como, entre otras, aplicar a los delincuentes de bajo riesgo encarcelados alternativas distintas a la prisión.

En cuarto lugar, la defensa alegó que COMPAS consideraba únicamente la edad de la persona en el momento de cometer el primer delito, la edad actual y los antecedentes penales y, además, no resultaba posible conocer cómo ponderaba cada factor porque Northpointe Inc. (la empresa creadora del sistema) no revelaba su contenido bajo el paraguas del secreto comercial y de empresa. Asimismo, la defensa alegó que COMPAS no se puedía evaluar y testar de forma objetiva sin conocer cómo funcionaba realmente, al existir una absoluta falta de transparencia.

En relación con ello, el Tribunal Supremo de Wisconsin adujo que la primera aseveración resultaba incorrecta, habida cuenta de que era público que COMPAS incluía 137 factores distintos clasificados en escalas de riesgo y necesidad que se dividían en cinco categorías: participación delictiva, relaciones y estilo de vida, personalidad y actitudes, familia y

exclusión social. Y respecto de ello, puso de manifiesto que el sistema generalmente tenía en cuenta, en la escala de riesgo de reincidencia básica, los cargos actuales del acusado, los cargos pendientes, el historial de detenciones pasadas, las incomparecencias previas al juicio, la estabilidad residencial, el estado laboral, los vínculos comunitarios y el abuso de sustancias; y en la escala de riesgo de reincidencia violenta, el historial de violencia, el historial de incumplimientos, los eventuales problemas educativos, la edad de la persona en el momento de la evaluación y la edad de la persona en el momento del primer arresto.

Respecto de la segunda alegación, el mencionado Tribunal adujo que tampoco resultaba cierta, habida cuenta de que COMPAS era un instrumento de evaluación de riesgos de cuarta generación que había sido probado objetivamente para determinar su confiabilidad y validez. Así, se expuso que, por un lado, el Centro Nacional de Tribunales Estatales (The National Center for State Courts) articulaba estándares para determinar si una herramienta de evaluación de riesgo era una buena herramienta y los resultados de COMPAS habían sido altamente satisfactorios; y, por otro lado, el Instituto Nacional de Tribunales de Drogas (National Drug Court Institute) había identificado a COMPAS como un instrumento recomendado, habiendo sido la única herramienta de evaluación de riesgos que recibió una calificación de "Bueno-Excelente" en términos de validez predictiva.

Así, el antedicho Tribunal concluyó que COMPAS era una herramienta confiable y válida para medir el riesgo de reincidencia y las necesidades del infractor y que, por ende, permitía a los tribunales crear sentencias más inteligentes, eficientes y personalizadas, con mejora del nivel de detección de aquellos delincuentes que deberían ir a prisión y aquellos que podrían beneficiarse, en cambio, de condenas alternativas.

Y, además, el Alto Tribunal de Wisconsin determinó que el derecho al debido proceso no requería la divulgación de las fórmulas utilizadas por los sistemas patentados para determinar el riesgo, bastando con que su uso fuera relevante para la consecución de los objetivos de la sentencia, con que la herramienta resultara confiable y válida, como ocurría en el caso de COMPAS (que había sido científicamente probada de forma independiente para comprobar su fiabilidad y validez) y con que no fuera el único factor determinante del fallo, debiendo siempre este estar apoyado en otros motivos y argumentos distintos de los resultados del sistema de IA (como ocurrió en el caso analizado, en que el tribunal de

instancia tuvo en cuenta principalmente la gravedad de los delitos y el nefasto historial penal del acusado).

Y, finalmente, la defensa alegó que COMPAS ignoraba y olvidaba las características individuales de cada ciudadano en favor de las características de grupo.

En relación con ello, el Tribunal Supremo de Wisconsin adujo que si bien la evaluación de riesgos realizada por el sistema colocaba al acusado en una determinada categoría de delincuentes, esta era individualiza para cada acusado, ya que recopilaba información de varias fuentes (entre otras, registros oficiales y entrevista con el mismo) que permitía generar un perfil completo, preciso e individual de cada uno.

No obstante, en la mencionada sentencia, con ánimo de poner de poner límite al uso de la herramienta COMPAS por parte de los tribunales sentenciadores, se fijaron las bases para el uso de la misma, especificando cómo debían presentarse en fase de instrucción los informes (PSI) que contuvieran evaluaciones de riesgo, que deberían contener las siguientes advertencias y cautelas escritas: que la naturaleza patentada de COMPAS impide revelar cómo se calculan los riesgos; que las evaluaciones de COMPAS no pueden identificar a individuos específicos de alto riesgo puesto estas se basan en datos grupales; que, aunque COMPAS se basa en una muestra de datos nacionales (EEUU), no ha existido ningún estudio de validación para la población de Wisconsin; que ciertos estudios han planteado preguntas sobre si las puntuaciones de COMPAS califican, de modo desproporcionado, a los delincuentes pertenecientes a minorías; y que COMPAS fue desarrollado específicamente para ayudar al "*Department of Corrections"* a tomar decisiones anteriores a la sentencia.

En relación con todo lo anterior, en el año 2016, un informe de ProPublica[505] determinó que los datos contenidos en el sistema COMPAS parecían sesgados en contra de las minorías, habida cuenta de que los acusados de raza negra tenían muchas más posibilidades que los acusados de raza blanca de ser juzgados por mayor riesgo de reincidencia de forma

[505] Véase Angwin, Larson, Mattu & Kirchner, 2016.

incorrecta, mientras que los acusados de raza blanca tenían más probabilidades que los de raza negra de ser calificados incorrectamente como de bajo riesgo.

La investigación que sirvió de base al mencionado informe consistió en el examen de más de diez mil casos de personas acusadas de cometer delitos en el Condado de Broward (Florida, EEUU), por ser una amplia jurisdicción que utilizaba la herramienta COMPAS en la toma de decisiones previas al juicio, y la comparación de las tasas de reincidencia previstas por esta (que distinguía entre "riesgo de reincidencia" y "riesgo de reincidencia violenta") con las tasas que finalmente resultaron reales durante un periodo de dos años.

Como consecuencia de ello, los investigadores descubrieron que mientras el sistema COMPAS predijo correctamente en un 61% de los casos la tasa de reincidencia delictiva, únicamente acertó en el 20% respecto de la tasa de reincidencia violenta. En relación con ello, el análisis reveló que:

-el sistema COMPAS a menudo predijo que los acusados negros tenían un mayor riesgo de reincidir del que realmente tenían, habiendo detectado que los que finalmente no reincidieron durante los dos años siguientes contaban, no obstante, con casi el doble de probabilidades de ser clasificados erróneamente como de mayor riesgo en comparación con los acusados blancos (45% frente a 23%);

-el sistema COMPAS a menudo predijo que los acusados blancos tenían un menor riesgo de reincidir del que realmente tenían, habiendo detectado que los que finalmente reincidieron dentro de los dos años siguientes habían sido calificados erróneamente como de bajo riesgo con casi el doble de frecuencia que los reincidentes negros (48% frente a 28%);

-con el sistema COMPAS, al hacer evaluaciones teniendo en cuenta antecedentes penales, reincidencia futura, edad y género, los acusados negros tenían un 45% más de probabilidades de que se les asignaran puntuaciones de riesgo de reincidencia más altas que los acusados blancos;

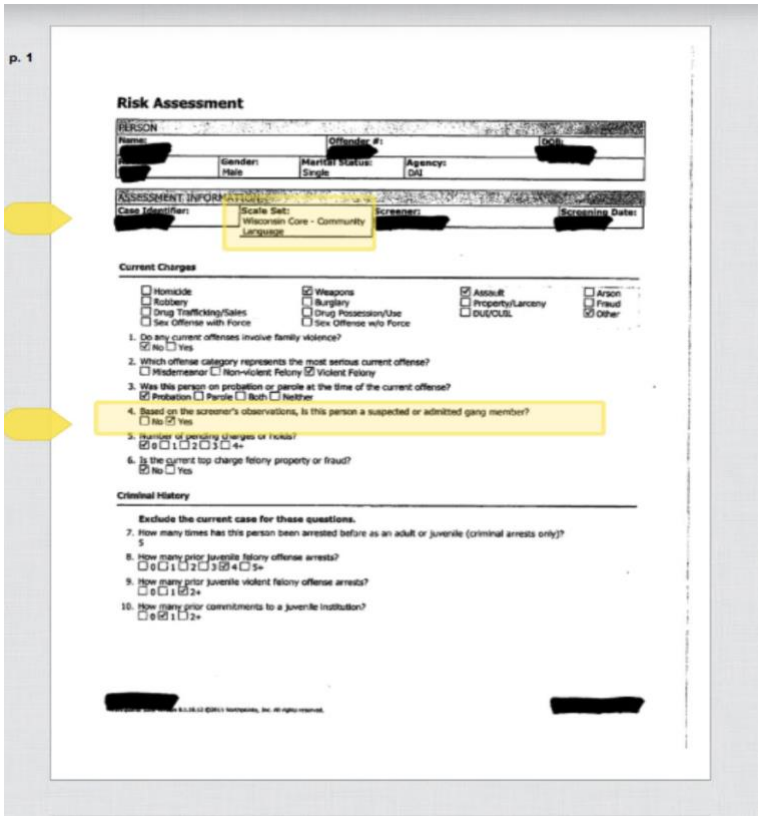
-con las predicciones del sistema COMPAS, los acusados negros tenían el doble de probabilidades que los acusados blancos de ser clasificados erróneamente como de mayor riesgo de reincidencia violenta y, asimismo, los reincidentes violentos blancos tenían un 63% más de probabilidades de haber sido clasificados erróneamente como de bajo riesgo de reincidencia violenta, en comparación con los reincidentes violentos negros; y

-con el sistema COMPAS, al hacer evaluaciones teniendo en cuenta antecedentes penales, reincidencia futura, edad y género, los acusados negros tenían un 77% más de probabilidades de que se les asignaran puntuaciones de riesgo de reincidencia violenta más altas que los acusados blancos.

Además, es interesante poner de manifiesto que los factores empleados por COMPAS, según se desprende del mencionado informe, están basados en un total de ciento treinta y siete cuestiones, contenidas en un formulario, con subfactores relativos a las siguientes quince áreas:

- Cargos actuales (seis preguntas)
- Antecedentes penales (dieciocho preguntas)
- Incumplimientos (seis preguntas)
- Criminalidad familiar (ocho preguntas)
- Compañías (seis preguntas)
- Abuso de sustancias (nueve preguntas)
- Residencia/Estabilidad (once preguntas)
- Entorno social (seis preguntas)
- Educación (nueve preguntas)
- Trabajo (quince preguntas)
- Ocio (ocho preguntas)
- Aislamiento social (nueve preguntas)
- Personalidad criminal (nueve preguntas)
- Ira (seis preguntas)
- Actitudes criminales (once preguntas)

Siendo la apariencia del antedicho formulario (primera página de ocho), la siguiente:


p. 1

Risk Assessment

PERSON

Gender: Male

Marital Status: Single

Agency: DAI

ASSESSMENT INFORMATION

Case Identifier:

Scale Set: Wisconsin Core - Community

Language

Screener:

Screening Date:

Current Charges

Homicide; Robbery; Drug Trafficking/Sales; Sex Offense with Force; Weapons; Burglary; Drug Possession/Use; Sex Offense w/o Force; Assault; Property/Larceny; DUI/OUIL; Arson; Fraud; Other

1. Do any current offenses involve family violence? No / Yes
2. Which offense category represents the most serious current offense? Misdemeanor / Non-Violent Felony / Violent Felony
3. Was this person on probation or parole at the time of the current offense? Probation / Parole / Both / Neither
4. Based on the screener's observations, is this person a suspected or admitted gang member? No / Yes
5. [illegible] 0 / 1 / 2 / 3 / 4+
6. Is the current top charge felony property or fraud? No / Yes

Criminal History

Exclude the current case for these questions.

7. How many times has this person been arrested before as an adult or juvenile (criminal arrests only)? 5
8. How many prior juvenile felony offense arrests? 0 / 1 / 2 / 3 / 4 / 5+
9. How many prior juvenile violent felony offense arrests? 0 / 1 / 2+
10. How many prior commitments to a juvenile institution? 0 / 1 / 2+

[redacted]

[redacted]

506

En relación con ello, es interesante hacer referencia a un estudio realizado en 2018 por Christopher Slobogin, Profesor de Derecho de la Universidad de Vanderbilt (Nashville, Tennessee, EEUU) y Megan Stevenson, Profesora de Derecho de la Universidad de Virginia (Virginia, EEUU), que aseguró que la edad tenía un peso de un 58% en la puntuación de riesgo de reincidencia violenta emitida por el sistema COMPAS[507], lo cual resulta, cuanto menos, sorprendente.

No obstante, tal y como se desprende de la guía práctica publicada por la compañía Equivant el 4 de abril de 2019[508], los factores incluidos en la ya mencionada (reciente y renovada) escala de Riesgo de Liberación Previa al Juicio-II ("*Pretrial Release Risk Scale II*") son más específicos y se enfocan únicamente en factores relevantes para determinar una posible falta de comparecencia y una nueva detención por la comisión de nuevos delitos graves en la fase previa al juicio (lo que demuestra un esfuerzo de la compañía por mejorar la objetividad), incluyendo:

[506] ProPublica, s.f..
[507] Véase Stevenson & Slobogin, 2018.
[508] Véase Equivant, 2019.

- Cargo más grave por delito mayor
- Casos pendientes
- Incomparecencias anteriores
- Detenciones previas (con fianza)
- Sentencias de prisión previas
- Historial de adicciones a sustancias estupefacientes
- Situación de empleo
- Duración de la residencia

Además de COMPAS, tal y como ya se ha anunciado con anterioridad, existen otras herramientas que emplean la IA para la evaluación de riesgos en la fase previa al juicio penal, siendo estas principalmente utilizadas en el ámbito de EEUU, donde cohabitan más de cientocincuenta modelos de las denominadas RATs ("*Risk Assessment Tools*") -a saber, entre otras, Domestic Violence (DV), Pretrial Risk Assessment Instrument (PTRA), Colorado Pretrial Assessment Tool (CPAT), Pretrial Release Risk Scale (PRRS), Delaware Pretrial Assessment Tool (DELPAT), Ontario Domestic Assault Risk Assessment Tool (ODARA), Minnesota Pretrial Assessment Tool (MNPAT), Ohio Risk Assessment System (ORAS), Level of Service/Case Management Inventory (LS/CMI), Pretrial Risk Assessment Information System (PRAISTX), Virginia Pretrial Risk Assessment Instrument (VPRAI) e Indiana Risk Assessment System (IRAS)-. Y es que los algoritmos se utilizan en los distintos sistemas judiciales de todo el país si bien, como se ha dicho, los sistemas específicos difieren según el estado (e, incluso, según el condado en algunos casos), siendo los principales (con adopción de sus propias versiones de cada uno de ellos según los casos) el ya analizado COMPAS, Public Safety Assessment (PSA) y Level of Service Inventory Revised (LSI-R).

Y es que, como expondré a continuación con más detenimiento, existen diferencias entre todos ellos, especialmente habida cuenta de que COMPAS, por un lado, tal y como se ha expuesto, evalúa para determinar el riesgo factores tales como la actividad delictiva, las relaciones y el estilo de vida, la personalidad y la conducta, la familia y la exclusión social de la persona investigada; por otro lado, el PSA únicamente tiene en cuenta criterios relacionados con la edad y la actividad e historia delictiva; y finalmente, el LSI-R, de forma intermedia, intercala datos relativos a la vida personal y criminal de la persona analizada.

Así, por un lado, el sistema Public Safety Assessment (PSA), es una herramienta (utilizada, entre otros, en el estado de Nueva Jersey, EEUU) que emplea la IA para la evaluación de riesgos, desarrollada en el seno del proyecto estadounidense "*Advancing Pretrial Policy and Research*" (APPR), liderado por el National Partnership for Pretrial Justice con el apoyo de Arnold Ventures[509], y lanzada al mercado de forma gratuita en el año 2015 por "*Laura and John Arnold Foundation*" (LJAF)[510]. Dicho sistema, a partir del uso de algoritmos, analiza nueve factores relativos a las personas investigadas que se hallan en una fase del proceso penal anterior al acto del juicio y predice el riesgo existente en relación a tres cuestiones: la probabilidad de incomparecencia ante el tribunal antes del juicio, la posibilidad de que la persona investigada sufra una nueva detención policial en caso de quedar en libertad en dicho periodo, y la posibilidad de que sufra una nueva detención policial por comisión de un delito violento en el mismo lapso de tiempo. Los mencionados nueve factores empleados por el sistema para evaluar el riesgo son: la edad del investigado en el momento de la detención, el eventual delito violento por el que haya sido detenido, los cargos pendientes en el momento del arresto, las condenas previas por delitos menores, las condenas previas por delitos graves, las condenas previas por delitos violentos, los incumplimientos de obligación de comparecer ante el tribunal en los últimos dos años, los incumplimientos de obligación de comparecer ante el tribunal con carácter previo a los dos últimos años, y las sentencias anteriores con condena a pena de prisión.

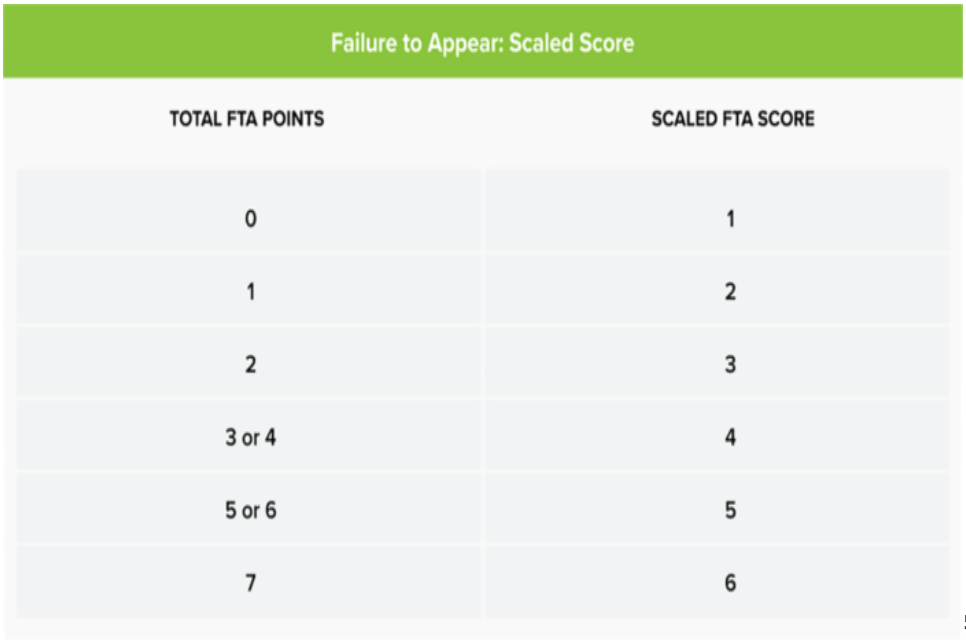
Tales factores, no obstante, no se tienen en consideración en su totalidad y de la misma forma para determinar los distintos posibles riesgos, según se dispone públicamente en la página web del proyecto "*Advancing Pretrial Policy and Research*" -APPR-. Así, en concreto, el sistema, para determinar el riesgo de incomparecencia ante el tribunal en la fase previa al juicio, tiene en cuenta los siguientes cuatro factores: cargos pendientes en el momento del arresto, condenas previas por delitos menores y/o por delitos graves, incumplimientos de la obligación de comparecer ante el tribunal en los últimos dos años, e incumplimientos de la obligación de comparecer ante el tribunal con carácter previo a los dos últimos años; para determinar el riesgo de comisión de nuevos delitos durante la fase previa al juicio, por su parte, se consideran siete factores: la edad del investigado en el

[509] Véase National Partnership for Pretrial Justice, s.f..
[510] Arnold Ventures, 2015.

momento de la detención, los cargos pendientes en el momento del arresto, las condenas previas por delitos menores, por delitos graves y/o por delitos violentos, los incumplimientos de la obligación de comparecer ante el tribunal en los últimos dos años y las sentencias anteriores con condena a pena de prisión; y, finalmente, para determinar el riesgo de comisión de delitos violentos durante la antedicha fase, se analizan los siguientes cinco factores: delito violento por el que, en su caso, ha sido detenido, cargos pendientes en el momento del arresto y condenas previas por delitos menores, por delitos graves y/o por delitos violentos.

Así, en un ejercicio de transparencia, se publican las tablas relativas a cada cuestión analizada, que muestran, en primer lugar, los puntos que se asignan a cada factor, y en segundo lugar, el nivel de riesgo que surge de la suma de todos ellos, en una escala del uno al seis. Y a modo de ejemplo, respecto del riesgo de incomparecencia ante el tribunal en la fase previa al juicio, se aplican las siguientes tablas:

Failure to Appear: Points		
PSA FACTOR	**RESPONSE**	**POINTS**
Pending charge at the time of the arrest	No	0
	Yes	1
Prior conviction (misdemeanor or felony)	No	0
	Yes	1
Prior failure to appear in the past 2 years	No	0
	Yes, just 1	2
	Yes, 2 or more	4
Prior failure to appear older than 2 years	No	0
	Yes	1



Failure to Appear: Scaled Score	
TOTAL FTA POINTS	SCALED FTA SCORE
0	1
1	2
2	3
3 or 4	4
5 or 6	5
7	6

[511]

De acuerdo con lo expuesto, el denominado PSA lleva a cabo una acción diferenciadora y más garantista que la del sistema COMPAS, habida cuenta de que hace públicos no solo los factores que tiene en cuenta para evaluar los riesgos, sino también la forma en que se combinan todos ellos para llegar a un resultado. Además, es importante poner de relieve que el mencionado sistema no requiere la realización de cuestionario alguno al presunto delincuente y, además, no tiene en cuenta, a diferencia de otras herramientas de evaluación de riesgos, características personales de los perfiles analizados (tales como el sexo, el lugar de residencia, los antecedentes familiares o la situación laboral). En relación con ello, en 2019 Arnold Ventures emitió su declaración de principios manifestando: "*La reforma de nuestros quebrados sistemas de justicia penal previa al juicio es una piedra angular del trabajo de justicia de Arnold Ventures. El sistema de justicia penal de Estados Unidos despoja a demasiadas personas de sus trabajos, familias, salud y dignidad; pone a las personas de color en riesgo, daña desproporcionadamente a los pobres, limita el potencial de los jóvenes atrapados en el sistema y no brinda a las personas afectadas las oportunidades que necesitan para volver al camino correcto. Y todo esto tiene un coste fiscal enorme. La justicia previa al juicio es una parte crítica y poco estudiada de este gran problema: en resumen, hay demasiadas personas en la cárcel que no deberían estar allí.* (…) *La evaluación de riesgos previa al juicio no es perfecta. Pero la toma de*

[511] Advancing Pretrial Policy & Research, s.f..

decisiones basada en datos es ciertamente menos sesgada que la intuición humana (…) *Además, la evaluación de riesgos siempre puede mejorar: los factores analizados y el peso de estos pueden y deben ser reevaluados y recalibrados a medida que los investigadores van aprendiendo, promoviendo así decisiones judiciales aún mejores y acercándonos a nuestro objetivo de eliminar la prisión preventiva injusta*. "[512]

No obstante, si bien es cierto que unos investigadores independientes evalúan rigurosamente, de forma contínua, el sistema PSA para maximizar su precisión y minimizar su impacto discriminatorio (lo cual es un avance significativo y una declaración de intenciones en toda regla), lo cierto es que no cuenta con una certificación pública de calidad (como sería deseable en los casos en que un sistema de IA se emplea por la Administración), ya que todo permanece en el ámbito privado.

Y, por otro lado, finalmente, procede hacer especial mención al sistema LSI-R (empleado, entre otros, en el estado de Idaho, EEUU) desarrollado por la empresa canadiense Multi-Health Systems, que analiza determinados factores de los presuntos delincuentes con el fin de asesorar a las autoridades fiscales y judiciales en la toma de decisiones relativas a su situación personal y/o, en su caso, tratamiento. Y es que esta herramienta tiene la especialidad, respecto de las otras dos analizadas, de que no solo mide o evalúa el riesgo de reincidencia de las personas investigadas sino también sus necesidades. Para ello, se examina y combina la información relativa a diez áreas principales de la vida de los investigados, tal y como se desprende del siguiente cuadro ilustrativo publicado por la empresa comercializadora del sistema, en un loable ejercicio de transparencia:

[513]

[512] Arnold Ventures, Statement of Principles on Pretrial Justice and Use of Pretrial Risk Assessment.
[513] Andrews & Bonta, s.f..

Tal y como se desprende de la página web de la mencionadan compañía, el sistema LSI-R cuenta con varias herramientas que tratan de maximizar sus niveles de transparencia y, en concreto, intentan fomentar la comprensión de cómo cada elemento analizado afecta al nivel de riesgo detectado.

De acuerdo con todo lo expuesto, es interesante advertir que, si bien en la última década numerosas organizaciones estadounidenses tales como Pretrial Justice Institute (PJI) han apostado con fuerza por la introducción en el país de herramientas de IA de evaluación de riesgos en la fase previa al juicio, recientemente han ido virando su posicionamiento.

Así, por ejemplo, por un lado, en 2018 la organización de derechos civiles The Leadership Education Fund[514], publicó un documento denominado "*A Shared Statement of Civil Rights Concerns*" que instó "*a las jurisdicciones a reconsiderar el uso de herramientas de evaluación de riesgos. Los instrumentos de evaluación de riesgos previos al juicio, aunque puedan parecer objetivos o neutrales, amenazan con intensificar aún más las diferencias injustificadas del sistema judicial y proporcionar un trato engañoso e inmerecido de parcialidad por parte una institución que necesita desesperadamente un cambio fundamental*." [515]

Por otro lado, en una encuesta realizada en el mismo año 2018 por la organización RTI International [516] a jueces, fiscales, abogados de la defensa y otros profesionales que intervienen en la fase penal previa al juicio en EEUU, el 82% de los abogados defensores manifestó su opinión de que el sistema de evaluación de riesgos PSA (Public Safety Assessment) contribuía a las disparidades raciales y étnicas en el sistema de justicia penal.[517]

[514] El "*Education Fund*" fue fundado en 1969 como rama de educación e investigación de "*The Leadership Conference on Civil and Human Rights*", la coalición de derechos civiles y humanos más antigua y más grande de EEUU, con más de doscientas organizaciones nacionales.
[515] Leadership Conference on Civil and Human Rights, 2018, pág. 1.
[516] Una organización sin ánimo de lucro con sede en Carolina del Norte (EEUU) que proporciona servicios técnicos y de investigación.
[517] DeMichele & otros, 2019.

Además, el 17 de julio de 2019, veintisiete prestigiosos investigadores de las universidades de Harvard, MIT, Princeton, NYU, UC Berkeley y Columbia, incluidos los directores de Berkman Klein Center, Martha Minow y Jonathan Zittrain, firmaron una declaración abierta en la que mostraron su preocupación por el uso de herramientas de evaluación de riesgos a través de la IA como medio para reducir las tasas de prisión provisional, concluyendo que: "*Las evaluaciones de riesgos previas al juicio no garantizan ni aumentan la probabilidad de mejores resultados previos al juicio. Las herramientas de evaluación de riesgos pueden simplemente cambiar u ocultar problemas de las actuales prácticas previas al juicio. Algunas jurisdicciones que han adoptado herramientas de evaluación de riesgos han visto tendencias positivas en los resultados previos al juicio, pero otras jurisdicciones han experimentado lo contrario. Dentro de las jurisdicciones que han logrado resultados positivos, no está claro si las herramientas de evaluación de riesgos fueron responsables de ese éxito o si ese éxito se debe a otras reformas o cambios que ocurrieron al mismo tiempo.* (…) (Este cuerpo técnico) *cuestiona la validez, ética y eficacia de las evaluaciones de riesgo previas al juicio. Por ejemplo, la mayoría de las herramientas de evaluación de riesgos son tecnología patentada y los acusados evaluados por estas herramientas no tienen la oportunidad de inspeccionar los algoritmos o sus datos subyacentes y rebatir los resultados.* (…) *Recomendamos encarecidamente recurrir a otras reformas*."[518]

Por su parte, en diciembre de 2019, la organización Community Justice Exchange publicó una guía con el fin de asesorar a los organizadores sobre cómo afrontar (y confrontar) el uso de las herramientas de evaluación de riesgos, como parte de parte de una estrategia más amplia para poner fin a los encarcelamientos masivos previos al juicio que acechan EEUU y a la excesiva supervisión. En dicha guía, en concreto, que se muestra muy crítica con los mencionados sistemas, se pone de manifiesto que "*la crisis de la prisión preventiva persiste tanto en lugares donde la evaluación de riesgos está consagrada en la toma de decisiones, como en lugares donde aún no se utiliza tal tipo de herramientas*"[519], y se pone como ejemplo al estado de Kentucky (EEUU), donde se implementó una herramienta de evaluación de riesgos obligatoria con la expresa intención de disminuir las cifras de prisión provisional, si bien los estudios han demostrado que no ha funcionado.

[518] Barabas & Benjamin, 2019, pág. 4.
[519] Pág. 5.

Y, finalmente, entre otros, el 7 febrero de 2020, el ya mencionado PJI emitió un informe en cuya virtud modificaba públicamente su postura, poniendo de manifiesto que "*Ahora vemos que las herramientas de evaluación de riesgos previas al juicio, diseñadas para predecir la posibilidad de comparecencia voluntaria de un individuo ante el tribunal sin ser detenido de nuevo, ya no pueden ser parte de nuestra solución para construir sistemas equitativos de justicia*"[520], y unos días más tarde, Madeline Carter y Alison Shameslos, directoras del ya anteriormente aludido proyecto APPR, emitieron un comunicado aclarando que "*Nuestros objetivos específicos son maximizar la comparecencia previa al juicio de los acusados, maximizar la libertad previa al juicio de estos y lograr un cambio sostenible a la vez que mantenemos la seguridad de la comunidad y trabajamos para eliminar las disparidades en las políticas y prácticas previas al juicio*", llamando a la colaboración de todos los agentes implicados y concluyendo que "*En última instancia, sin importar las herramientas o la estrategia, sin un compromiso subyacente con la justicia y la equidad, las distintas jurisdicciones tendrán dificultades para lograr sus objetivos. APPR cree que el uso de nuestra herramienta de evaluación previa al juicio es coherente con los principios de nuestro proyecto y puede contribuir a lograr nuestros objetivos, pero solo si se usa de manera responsable, como parte de un esfuerzo integral y para los propósitos establecidos, que no son otros que los de lograr justicia y equidad*." [521]

En relación con lo expuesto, procede poner de manifiesto que, a partir de los resultados de diversas investigaciones, es sabido que tras una indudable apariencia de seriedad y rigurosidad, las herramientas de evaluación de riesgos generalmente emplean tres tipos de datos distintos:

-*datos* "*estáticos*", que no cambian con el tiempo, tales como la edad de una persona en el primer arresto o los antecedentes penales, entre otros, y que a menudo se obtienen del expediente de una persona, no de una entrevista personal;

-*datos* "*dinámicos*", que sí que varían con el tiempo, tales como la situación de empleo y residencia, el consumo de sustancias estupefacientes o el uso de drogas, entre otros, y suelen recopilarse a través de una entrevista personal; y

[520] Pretrial Justice Institute, 2020, pág. 1.
[521] Advancing Pretrial Policy & Research, 2020.

-*datos* "*subjetivos*", que comprenden información sobre la persona investigada que, a través de un entrevistador o un oficial judicial, el sistema registra y, por ende, se basan en una interpretación subjetiva. Entre ellos pueden incluirse evaluaciones sobre la conducta, las relaciones con los demás o el estado de ánimo.[522]

No obstante, tales datos tenidos en cuenta por los algoritmos, tal y como ya se ha advertido con anterioridad, pueden contener sesgos o resultar incompletos. Así, si las informaciones se emplean bajo la creencia de que son objetivas y completas, existe la posibilidad de que arrojen resultados que no se ajusten a la realidad, ya que, entre otras, ¿cómo puede predecirse realmente el riesgo de que una persona sufra futuras detenciones sin tener en cuenta el sesgo existente en los antecedentes policiales? ¿cómo puede evaluarse el riesgo de incomparecencia ante el Tribunal sin tener en cuenta las condiciones personales y estructurales que hicieron que una persona incompareciera en el pasado?, etc.

Si bien es cierto que, cada vez más, las compañías que diseñan y comercializan los programas de IA de evaluación de riesgos están tendiendo a eliminar de sus algoritmos factores estrictamente personales de los individuos a analizar, tales como el sexo, la raza o la edad, lo cierto es que, como ya se ha visto, muchos de ellos todavía siguen considerándolos (de forma directa o indirecta) y, en muchas ocasiones, de forma opaca.

En España, actualmente, no contamos con ningún instrumento de IA que asista al Ministerio Público y a los jueces en la fase de instrucción del proceso penal (pero sí en la fase de ejecución de las penas, como ya se ha expuesto con anterioridad). No obstante, lo que está claro es que, desde luego, su uso en los términos previstos por el Tribunal Supremo de Wisconsin (y por la gran mayoría de estados y condados estadounidenses), podría resultar no solo ilegal sino absoluta y frontalmente inconstitucional. Y es que ello, por un lado, podría vulnerar el derecho a la igualdad y a la no discriminación regulado en los artículos 20 a 23 de la Carta de Derechos Fundamentales de la UE y en el 14 de nuestra Constitución Española; y, por otro lado, podría vulnerar el derecho a la defensa regulado en los artículos 48 de la Carta de Derechos Fundamentales de la UE y el en artículo 24 de nuestra Constitución Española.

[522] Desmarais & Loewder, 2019, pág. 36.

En primer lugar, debe hacerse referencia a la posible colisión en el ámbito europeo y, en concreto, en España, entre la toma en consideración por los algoritmos de características eminentemente personales de los invididuos analizados, y el derecho a la igualdad y a la no discriminación previsto en el artículo 14 de la Constitución Española.

En concreto, tal último mencionado precepto dispone: "*Los españoles son iguales ante la ley, sin que pueda prevalecer discriminación alguna por razón de nacimiento, raza, sexo, religión, opinión o cualquier otra condición o circunstancia personal o social*."

Y es que la igualdad es, sin duda alguna, uno de los pilares fundamentales de cualquier Estado de Derecho y, de hecho en nuestra Carta Magna viene configurado no solo como derecho fundamental *stricto sensu*, sino también como valor superior del ordenamiento jurídico en el artículo 1.1, que dispone: "*España se constituye en un Estado social y democrático de Derecho, que propugna como valores superiores de su ordenamiento jurídico la libertad, la justicia, la igualdad y el pluralismo político*."

Al respecto, el Tribunal Constitucional ha puesto de manifiesto que el derecho a la igualdad es un "*derecho fundamental carente de autonomía propia, en cuanto se da solo en relación con otros derechos, a los que, por decirlo así, modula, de acuerdo con la igualdad entendida como valor y proclamada en el artículo 1.1 de la Constitución*".[523] Y es que resulta evidente que el derecho a la igualdad siempre se propugna en relación con otros derechos en juego, como la libertad, el aceso a la justicia, el honor o la propiedad, por ejemplo. Es decir, se entiende que existe igualdad siempre y cuando dos o más individuos, en iguales circunstancias, puedan tener acceso de forma idéntica a los distintos derechos que les reconoce el ordenamiento jurídico, salvo en aquellos casos en que exista una justificación objetiva y razonable para dar un trato desigual (lo cual es una manifestación más, justamente, de ese principio de igualdad).

Y es que en relación con lo anterior, Andrés Ollero Tassara, Catedrático de Filosofía del Derecho de la Universidad de Granada, establece: "*No nos hallamos, pues, ante un mandato positivo de trato uniforme, las mismas exigencias de igualdad obligarán,*

[523] Auto del TC nº862/1986, de 29 de octubre.

paradójicamente, a tratar de manera desigual a aquellos ciudadanos que se encuentran en situación de relevante diversidad"[524], ya que, como dispone el Tribunal Constitucional: "*El principio de igualdad no prohíbe que el legislador contemple la necesidad o conveniencia de diferenciar situaciones distintas y de darles un tratamiento diverso, que puede incluso venir exigido, en un Estado social y democrático de Derecho, por la efectividad de los valores que la Constitución consagra con el carácter de superiores del ordenamiento*".[525] Así configurado el derecho a la igualdad, en palabras del mencionado catedrático: "*Se entiende por discriminación una desigualdad de trato carente de justificación objetiva y razonable, que ha de apreciarse en relación a la finalidad y efectos de la medida, cuidándose la adecuada relación de proporcionalidad de los medios empleados a tal efecto.*"[526]

De acuerdo con lo expuesto, procede analizar si el uso de factores eminentemente personales o sociales y no estrictamente relacionados con los historiales delictivos de las personas investigadas por parte de los sistemas de IA con el fin de evaluar su riesgo de incomparecencia ante el tribunal (riesgo de fuga) o de cometer nuevos delitos durante la fase previa al juicio (reincidencia), resultaría o no constitucional y legal. En cualquier caso, hay que tener en cuenta que el uso de las herramientas analizadas debería quedar sujeto en la actualidad al cumplimiento de lo dispuesto en la LO 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, en los términos ya expuestos en relación con los sistemas policiales de evaluación de riesgos y, en su caso, cuando se apruebe y publique el texto definitivo, al cumplimiento de lo previsto en el Reglamento de IA, en los mismos términos. [527]

En relación con ello, y a modo de prevención, por un lado, el ya mencionado artículo 14.2 de la LO 7/21, de 26 de mayo, dispone que las decisiones basadas únicamente en un tratamiento de datos automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos negativos para el interesado o que le afecten significativamente (como es

[524] Ollero, 1992, pág. 546.
[525] Sentencia del TC nº34/1981, de 10 de noviembre.
[526] Ollero, 1992, pág. 546.
[527] Véanse páginas 209-214.

el caso) no se basarán en las categorías especiales de datos personales contempladas en el artículo 13 (es decir, aquellos que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas o la afiliación sindical, así como los datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, y los datos relativos a la salud o a la vida sexual o a la orientación sexual de una persona física), salvo que se hayan tomado las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado. Y, por su parte, el apartado 3 de tal precepto prohíbe la elaboración de perfiles que dé lugar a una discriminación de las personas físicas sobre la base de categorías especiales de datos personales establecidas en el mencionado artículo 13, lo que, sin duda, constituye una garantía directa y clara del derecho a la igualdad y a la no discriminación en relación con el uso de tales herramientas en nuestro país (que, no obstante, todavía no están habilitadas legalmente).

Y, por otro lado, el artículo 13 del citado cuerpo legal, tal y como ya se expuso en páginas anteriores[528], respecto del resto de casos, prohíbe el tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas o la afiliación sindical, así como el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, los datos relativos a la salud o a la vida sexual o a la orientación sexual de una persona física, salvo que este sea estrictamente necesario, con sujeción a las garantías adecuadas para los derechos y libertades del interesado y cuando, o bien se halle previsto por una norma con rango de ley o por el Derecho de la Unión Europea; o bien resulte necesario para proteger los intereses vitales, así como los derechos y libertades fundamentales del interesado o de otra persona física; o bien se refiera a datos que el interesado haya hecho manifiestamente públicos.

En relación con ello, no obstante, me surgen varias preguntas: ¿resultaría contrario en todo caso al derecho a la igualdad y a la no discriminación tener en cuenta para evaluar los mencionados riesgos la nacionalidad de una persona? ¿Y el sexo? ¿Y la edad? ¿Qué pasa con la religión? ¿Y con la situación económica? En mi opinión, si bien, *per se*, la consideración de tales factores para tomar una decisión relativa a la persona, que puede

[528] Véase página 212.

incidir en derechos tan relevantes como su libertad, puede antojarse discriminatoria, entiendo que, no obstante, podría resultar legalmente viable siempre y cuando se cumpliera las condiciones antedichas.

Bajo mi punto de vista, no obstante, para poder efectuar un tratamiento legítimo de las mencionadas categorías especiales de datos personales, no solo bastaría con que se cumplieran los requisitos previstos en la LO 7/21, de 26 de mayo, sino que, además, debería subyacer el cumplimiento del requisito estrella marcado por el Tribunal Constitucional: la existencia de una justificación objetiva y razonable.

Ante ello, no obstante, surge la siguiente cuestión: ¿qué es una justificación o fundamentación objetiva y razonable? En mi opinión, por un lado, tal expresión lleva implícita, sin duda, la necesidad de que el poder público que se vea involucrado en la elaboración (igualdad en la ley -legislador-) o aplicación (igualdad ante la ley -operador jurídico-) de una norma jurídica que prevea ciertos rasgos personales o sociales de los ciudadanos para diferenciarlos en el trato, explique o exponga de forma razonada y pública los motivos por los cuales tal distinción se lleva a cabo; y, por otro lado, entiendo que la mencionada justificación debe atender a criterios que tengan una base objetiva (es decir, sin subjetividades y respaldados por evidencias empíricas, científicas, etc), y razonable, lo cual deberá determinarse en cada caso, debiendo hacer los poderes públicos -obligados a promover las condiciones para que la igualdad del individuo y de los grupos en que se integra sean reales y efectivas (artículo 9.3 de la Constitución Española)- un balance entre los derechos en juego, las circunstancias y las finalidades que rodean al trato aparentemente "desigual", que debe resultar proporcionado y necesario.

En el caso de las herramientas de evaluación de riesgos, no obstante, nos hallamos ante un caso especialmente peculiar y peligroso.

Y es que, por un lado, hay datos personales que avalan de forma objetiva (o al menos, eso opino yo) un trato desigual, como por ejemplo el sexo, la edad o la adicción al alcohol o a las sustancias estupefacientes. Así, resulta evidente y objetivo que, según se desprende de las estadísticas, en 2018, en España, fueron condenados por cometer un delito ciento setenta y un mil quinientos treinta y seis hombres frente a cuarenta y ocho mil setecientas

cuarenta y ocho mujeres, y por cometer dos delitos treinta y cinco mil seiscientos cuarenta y cinco hombres, frente a tan solo siete mil ochenta y nueve mujeres[529]; y que, en 2019, fueron condenados por sentencia firme en nuestro país doscientos ochenta y seis mil novecientos treinta y un adultos y catorce mil ciento doce menores.[530] Asimismo, en el Informe sobre Drogas del año 2019 publicado por el Gobierno de España[531] se hizo referencia a una encuesta del año 2016 que indicó que alrededor del 40% de las personas internas en un centro penitenciario había consumido cannabis durante los treinta días anteriores a su ingreso en prisión (la cocaína y la heroína fueron las siguientes drogas más usadas), presentando la mayoría de los encuestados un patrón claro de policonsumo anterior a su entrada en la cárcel.

La cuestión, no obstante, es la siguiente: ¿qué hay detrás de esas estadísticas? En este caso, datos relativos a las condenas penales firmes recaídas en un periodo de tiempo determinado sobre un grupo de personas diferenciado por su sexo, por su edad y por su adicción a sustancias tóxicas, lo cual entiendo que resulta estrictamente objetivo. Sin embargo, inevitablemente surge la pregunta: ¿lleva ello sesgos implícitos? ¿llegan a los tribunales más casos con acusados hombres, adultos y adictos a sustancias estupefacientes como consecuencia de prácticas policiales cuestionables? ¿tienden los tribunales, por convicciones o motivos espurios, a condenar más a los hombres que a las mujeres, a los adultos que a los menores o a los toxicómanos que a los no adictos? ¿Existen prejuicios en ese sentido que no hacen más que generar datos sesgados sobre los que elaborar estadísticas que luego pueden servir para justificar que un algoritmo de un trato desigual? Realmente es algo muy difícil de adivinar, si bien en este caso, considero que las estadísticas reflejan de forma bastante fiel la realidad (siempre salvo prueba en contrario).

No obstante, hay que poner de manifiesto que existen rasgos de la personalidad que históricamente, en el ámbito criminal, han ido tristemente ligados a la discriminación, lo cual se ha ido poniendo de manifiesto a través de múltiples investigaciones, denuncias y estudios académicos y científicos. Así, la raza, la etnia, la nacionalidad, el lugar de

[529] Fernandez , 2019.
[530] Instituto Nacional de Estadística, 2020.
[531] Gobierno de España, 2019, pág. 30.

residencia, la tendencia política o la situación económica (a pesar de que esta última información no es considerada categoría especial de datos personales) han determinado en numerosas ocasiones la dispensa de un trato penal desigual a los ciudadanos de un modo injusto e infundado. Y es que, por ejemplo, en la ciudad estadounidense de San Francisco, aunque la población afroamericana representaba tan solo alrededor del 6% del total, entre los años 2008 y 2014 llegó a representar el 43% de la población penitenciaria[532] ; y, asimismo, si bien la población hispana en EEUU en el año 2020 representaba el 18,5% de la población total[533], suponía el 30,4% de la población de las prisiones del país.[534] En 2017, por su parte, por ejemplo, además, los senadores Kamala Harris (California, EEUU) y Rand Paul (Kentucky, EEUU) introdujeron la "*Pretrial Integrity and Safety Act of 2017*" que proponía sustituir, en fase de instrucción, la fianza económica para eludir la prisión provisional por los informes de evaluación de riesgo, con el fin de que la libertad anterior al juicio se basara en el peligro del individuo en vez de en su riqueza, puesto que lo contrario obligaba a ir a prisión a aquellos con menos posibilidades económicas.[535]

En relación con los datos estadísticos relativos a tales minorías, pues, a diferencia de lo que ocurre con informaciones presuntamente más objetivas, entiendo que el tratamiento debe ser especialmente cauteloso, ya que difícilmente podrán sustentar una explicación o justificación objetiva y razonable si sirven de elementos de trato diferenciador, puesto que más bien al contrario, su uso puede implicar una perpetuación de los segos históricamente existentes e indebidamente tolerados.

En tal sentido, bajo mi punto de vista, pues, no debe darse un tratamiento homogéneo a todos los rasgos de la personalidad en relación con los efectos legales que puede ocasionar su inclusión en los sistemas de evaluación de riesgos, puesto que, como se ha visto, algunos de ellos tienen un sustento subjetivo y sesgado, y otros (en principio) no. De acuerdo con ello, con carácter general, yo me muestro partidaria de que los algoritmos tengan en cuenta factores tales como la edad y el sexo, pero me muestro reticente a que se incluyan la raza,

[532] Williams, 2019.
[533] Gobierno de Estados Unidos, 2020.
[534] Gobierno de Estados Unidos, 2021.
[535] Green, 2020, pág. 2.

la nacionalidad, la etnia, la religión, la opinión y demás rasgos personales o sociales del individuo, ya que cuentan con una base que difícilmente puede considerarse objetiva.

De todas formas, resulta claro que esta no es una cuestión nada fácil de solventar, puesto que puede variar mucho según el caso concreto, por lo que entiendo que siempre debería analizarse de forma concreta si puede existir o no una justificación objetiva y razonable.

Ello no obstante, en mi opinión, y a los efectos de evitar problemas legales y dotar a los ciudadanos de una mayor seguridad jurídica, lo deseable sería que se regulara de forma expresa el uso de esta clase de herramientas de IA, en relación con lo dispuesto en los artículos 13 y 14 de la LO 7/21 de 26 de mayo, con una clara referencia a los factores o criterios que dichos sistemas deberían tener en cuenta para efectuar las predicciones y qué peso debería otorgarse a cada uno de ellos en cada caso, pudiendo así los operadores jurídicos acudir siempre a una guía de criterios legalmente prevista para tomar sus decisiones. Y es que, si bien los diseñadores de dichos sistemas aseguran que los algoritmos pueden ser objetivos, la objetividad, según entiendo, es un concepto muy relativo, especialmente en el ámbito jurídico, puesto que deberá ser el poder legislativo quien decida, en cada momento, qué considera objetivo y qué no, y quien elija, asimismo, qué elementos o factores deben incluirse en "la máquina" que asista al juez en su toma de decisiones y disponga qué resultados deben equivaler a niveles de riesgo bajo, moderado o alto, lo cual deberá ser ejecutado, con garantías de calidad, por las empresas que la diseñen, fabriquen y comercialicen (una vez más, pasar el filtro de una Agencia de certificación de calidad de sistemas de IA como la que se viene reclamando a lo largo de la presente tesis doctoral sería una muy buena solución). Y es que, nos guste o no, hay una realidad: nada en la ciencia queda libre de la influencia humana y cultural, puesto que de lo contrario, nos acabaría dominando.

Así, por ejemplo, por un lado, imaginemos una herramienta de IA de evaluación de riesgos que se emplee únicamente para determinar el peligro de que una persona detenida por la comisión de delito terrorista (radicales religiosos o políticos) pueda cometer nuevos actos de la misma índole. ¿Acaso no serían la religión y la eventual militancia en un partido político concreto factores a tener en cuenta? ¿No es una realidad que (me atrevo a decir) casi el 100% de los extremistas religiosos profesan una determinada religión que les

empuja a su lucha y casi el 100% de los terroristas políticos han tenido vínculos con determinados partidos? Resulta una obviedad que, en tal sentido, la religión y la tendencia política de un individuo son elementos importantes a tener en cuenta y pueden aportar información valiosísima en la evaluación de sus riesgos y, por ende, considero que puede existir una justificación objetiva y razonable para incluirlos como factores a analizar en las bases de datos empleadas por los algoritmos.

Imaginemos, por otro lado, una herramienta de IA diseñada para la evaluación del riesgo de fuga. ¿No es patente que un extranjero sin residencia en España suele tener más riesgo de incomparecencia que un nacional con residencia en España? Resulta evidente que lo que tiene que valorarse en cada caso para evaluar el riesgo de fuga de una persona es, entre otros, el arraigo existente, pero no veo mal que el algoritmo, en tales casos y con tal fin, tenga en cuenta la nacionalidad y el lugar de residencia del individuo para asesorar sobre el nivel de peligro. No obstante, cierto es que los potenciales efectos discriminatorios podrían mitigarse si, en vez de incluir la nacionalidad en la base de datos, únicamente se valorara si la persona es nacional o extranjera y si tiene su residencia en España o fuera del país, para que el algoritmo pudiera hacer una correcta combinación de informaciones sin entrar en demasiado detalle. En relación con ello, es interesante poner de manifiesto que el Instituto Nacional de Estadística elabora y publica unas estadísticas anuales por tipo de delito y nacionalidad de sus autores, si bien únicamente distingue entre España, Estados Miembros de la UE, países europeos no incluidos en la UE, Asia, África, América y Oceanía, lo cual podría llegar a servir de base para los fines expuestos.

No obstante, siendo yo la primera que entiende que debemos aunar esfuerzos para, de una vez por todas, acabar con la multitud de sesgos y prejuicios que acechan nuestras comunidades, también quiero hacer una pequeña puntualización. Y es que tengo la sensación de que, en ocasiones, especialmente en la Unión Europea (donde el máximo respeto a las libertades y a los derechos fundamentales es el pilar del sistema jurídico, como debe ser), nos perdemos en el purismo, la literalidad o la "exquisitez jurídica", lo que a veces impide que tomemos decisiones eficientes y beneficiosas para todos.

Así, por ejemplo, cierto es que la nacionalidad, por concepto, no debería ser un factor a tener en cuenta para evaluar el riesgo de fuga o reincidencia de una persona, pero en

determinadas ocasiones entiendo que considerarlo podría ser muy útil, puesto que aportaría muchísima y muy útil información.

En relación con ello, por un lado, los jueces de instrucción (así como los policías y los fiscales) saben que (al menos en Cataluña) a lo largo de su carrera se han topado con múltiples personas de una determinada nacionalidad que se dedican a cometer delitos de robo con fuerza en interior de vivienda, lo cual llama poderosamente la atención, puesto que los delincuentes de tal nacionalidad no tienen incidencia apenas en otros tipos de delito. Además, también es sabido, por la máxima de la experiencia, que el patrón delictivo de los nacionales de tal concreto país que cometen dicha clase de ilícitos penales suele ser exactamente el mismo: personas que tienen Letrado particular, se acogen a su derecho a no declarar (en ocasiones, como mucho, en las comparecencias de prisión reconocen tener a la familia en su país de origen), necesitan intérprete porque no dominan el idioma español, en caso de quedar libres suelen no comparecer ante la justicia, se muestran conformes con las expulsiones, y por las informaciones que ha venido manejando la policía a lo largo de los años, en la mayoría de casos pertenecen a mafias que les contratan simplemente para venir a nuestro país, cometer el máximo número de robos posible en un tiempo determinado y luego volver a su lugar de origen. Ello, inevitablemente, en los casos en que se instruyen hechos compatibles con tales tipo delictivos y las personas investigadas son de dicha nacionalidad, lleva a las autoridades policiales, fiscales y judiciales a tener en cuenta tal extremo para decidir sobre su situación personal. Y es que, lamentablemente, cierto es que a lo mejor una persona nacional de tal país tiene el mismo riesgo de fuga y de reiteración delictiva que un ciudadano de otra nacionalidad, pero es una realidad que los primeros suelen tener un *modus operandi* que no resulta compatible con su puesta en libertad. Bajo mi punto de vista, en estos casos la consideración de la nacionalidad concreta podría llegar a quedar justificada por una explicación objetiva (no habría más que realizar y analizar estadísticas de aquellas personas de cierta nacionalidad que cometen este tipo de delitos, no comparecen ante el tribunal en caso de quedar en libertad y, si lo hacen, resultan condenadas y son expulsadas, por ejemplo, que deberían actualizarse constantemente) y razonable, si bien tras conversaciones con compañeros y juristas, he advertido que hay una buena parte de la doctrina que está en contra de tal opinión, por entender que en ningún caso es respetuosa con el derecho a la igualdad y a la no discriminación.

Por otro lado, los policías, fiscales y jueces de instrucción, especialmente los que lidian con delitos relativos a la violencia doméstica y de género, por su experiencia, saben que los nacionales de ciertos países, por su cultura y forma de relacionarse con la familia, son más propensos a cometer tal tipo de delitos, incluso de forma reincidente, que otros. Así, por ejemplo, es llamativo el elevado número de familias de determinados países que se hallan sometidas en bucles de violencia doméstica y tienen un patrón que se repite constantemente: la esposa, que recibe malos tratos por parte del marido (que en la mayoría de ocasiones, se extienden a los hijos comunes), no presenta denuncia por miedo a las represalias del esposo y de su propia familia (que en muchos casos es conocedora y consentidora de la situación), ya que el divorcio en tales países suele ser sinónimo de repudia, y aguanta hasta que un vecino, un conocido o la propia policía descubre lo que está sucediendo y procede a ponerlo en conocimiento de las autoridades competentes. Tras ello, si bien en muchas ocasiones las víctimas se acogen a su derecho a no declarar contra sus maridos o familiares en sede judicial, los procedimientos suelen seguir adelante por las declaraciones de los testigos y los informes del Médico Forense y, en aquellos casos en que se concede una orden de alejamiento, es más que probable que la mujer vuelva a los pocos días (si no al día siguiente) a solicitar su retirada, por las fatales consecuencias familiares y económicas que su vigencia conlleva. En tales casos, para evaluar el riesgo de reincidencia del marido, ¿acaso no sería conveniente que los algoritmos tuvieran en cuenta la nacionalidad, a sabiendas del calado cultural que existe en ciertos países sobre las relaciones del hombre con su mujer y sus hijos? ¿de verdad creemos que tal dato no influye, de forma consciente o inconsciente, en la toma de decisiones de los jueces a la hora de valorar el riesgo de reincidencia para acordar una orden de alejamiento o un ingreso en prisión? ¿no sería oportuno elaborar estadísticas sobre el grado de reincidencia que tienen los autores de delitos de violencia doméstica y de género de determinadas nacionalidades para poder así tener una base objetiva sobre la que respaldarse? Yo estoy convencida de que sí.

Y, finalmente, también es sabido por los distintos profesionales que intervienen en la fase de instrucción, que cuando existen niveles de pobreza extrema, es más probable que exista riesgo de reincidencia en el caso de delitos contra la propiedad, al igual que cuando existen grandes niveles de riqueza, resulta más previsible la existencia de un elevado riesgo de fuga. Y ello es una cuestión demostrada por la experiencia de casos pasados que, en mi

opinión, debería tenerse en cuenta, de forma acotada y limitada, desde luego, por los algoritmos que miden el riesgo de las personas investigadas a través de la IA.

Cierto es que no resulta ni agradable ni desde luego políticamente correcto señalar a los ciudadanos de un país o de una situación económica determinada, de forma general, en relación con algo tan grave como lo expuesto, pero la realidad es que la experiencia demuestra, al menos en mi caso, que existe un patrón de conducta bastante marcado que resulta muy poco responsable obviar y que, bajo mi punto de vista, debería poder servir de sustento para otorgar una explicación objetiva y razonable que justificara la inclusión de ciertos factores personales (no solo de "historia criminal" de cada uno) para ser tenidos en cuenta por los sistemas de IA que evalúan riesgos, con el fin de mejorar su nivel de éxito y precisión.

De lo que me he dado cuenta, no obstante, en los anteriormente mencionados intercambios de opiniones con compañeros y demás profesionales del ámbito jurídico, es que normalmente las personas que se muestran disconformes con la consideración de la nacionalidad, el lugar de residencia o la situación económica concretos de un individuo en la evaluación del riesgo de fuga o de reincidencia, argumentan su postura alegando que lo que debe tenerse en cuenta es el caso concreto, de forma individualizada, y que no podemos basarnos en datos generales y estadísticos ajenos para atribuirle un cierto nivel de riesgo a un individuo determinado. Así, entienden que lo contrario implicaría, no solo ir en contra de la indivudualización y de la presunción de inocencia, sino también fomentar la perpetuación de sesgos pasados. Y es que, por ejemplo, en caso de entender que el hecho de que una persona tenga menos recursos y se halle en situación de desempleo (al menos, oficial) debe ponderarse por el algoritmo para predecir su riesgo de cometer nuevos delitos contra la propiedad, se considera que supone un modo de discriminación indirecta, puesto que, por ejemplo, las comunidades negras, de inmigrantes y de personas pobres a menudo son excluidas de forma sistemática de aquellos trabajos con salarios dignos que se ofrecen en la economía oficial, por lo que tienen trabajos de economía sumergida y con bajos sueldos; y, asimismo, en caso de considerar que el hecho de que una persona tenga inestabilidad domiciliaria debe ser tenido en cuenta para determinar su riesgo de incomparecencia o de reiteración delictiva de forma algorítmica, se entiende asimismo que es un modo de discriminación indirecta, puesto que las comunidades con altas tasas de

desahucio e inestabilidad de la vivienda suelen estar excesivamente vigiladas y criminalizadas.

Ello, por un lado, me parece muy acertado, y por eso, desde luego, soy partidaria de que la última palabra en la toma de decisiones la tenga siempre un humano, especialmente en supuestos como el analizado, que pueden afectar seriamente a un derecho fundamental tan sensible como es el de la libertad. Pero, por otro lado, me parece muy *naïf* o demasiado conservador tal pensamiento, ya que es evidente que los jueces, desde siempre, han tenido en cuenta la realidad de los casos pasados similares o muy parecidos para tomar decisiones relativas a individuos concretos (de hecho, esa es la base del valor de la jurisprudencia, que según el artículo 1.6 del Código Civil español no es fuente del Derecho pero debe complementar el ordenamiento jurídico). Lo que sucede, no obstante, es que la comparación con casos pasados puramente dicha no es algo que se plasme normalmente en los Autos de medidas cautelares, ya que suele quedarse en ese (al menos por el momento) inaccesible *black box* del juez (y es que difícilmente un instructor argumentará en su resolución que como sabe, por su experiencia, que un gran número de personas de una determinada nacionalidad suelen estar contratadas por mafias para cometer robos con fuerza en interior de viviendas y tienen vuelo de vuelta a su paías ya reservado y pagado, entiende que el concreto investigado tiene riesgo de fuga, aunque en realidad lo piense). Y, en cualquier caso, no cabe olvidar que muchos de los factores cuya inclusión en los sistemas de IA de evaluación de riesgos ahora se cuestiona, están ya a día de hoy previstos por la propia legislación para guiar y asistir al juez instructor en su toma de decisiones, tal y como se verá más adelante, por lo que quizás lo que habría que plantearse es una reforma de más calado en este ámbito y, así, tal y como apunta Ben Green, doctorando de la Universidad de Harvard (Massachusetts, EEUU), tras el estudio en profundidad de los sistemas de IA de evaluación de riesgos: "(…) *las evaluaciones de riesgo pueden reinterpretarse para apuntar hacia una reforma de la justicia penal más sustantiva. Un desafío adecuado a las evaluaciones de riesgos no requiere reformas técnicas o de procedimiento, sino una "reforma profunda" que proporcione una nueva interpretación tanto de las evaluaciones de riesgo como del sistema de justicia penal.*"[536]

[536] Green, 2020, pág. 2.

Y es que entiendo que los sistemas de IA deben ir dotados de un plus de transparencia, no solo para garantizar el derecho a la defensa y otros relacionados con el mismo, sino también para asegurar un igual tratamiento jurídico a los millones de ciudadanos sobre los que se puede aplicar, siendo que la incidencia en las vidas de las personas de un sistema de tal índole es exponencialmente infinitamente mayor a la que puede tener un solo juez.

Además, considero que en la configuración de las bases de datos empleadas por tales herramientas de IA debe imperar como elemento diferenciador e inflexible la exigencia de que se nutran única y exclusivamente de datos relativos a condenas firmes o hechos objetivos y completos, ya que es la única forma de asegurarse (al menos, mínimamente) de que tales informaciones han pasado por un filtro judicial.

Así, por un lado, a mi entender, no podrían resultar válidas, por colisionar con el derecho a la presunción de inocencia de los ciudadanos, las informaciones relativas a antecedentes policiales o detenciones previas, y tampoco las referentes a antecedentes judiciales que no hayan terminado por resolución firme; y, por otro lado, bajo mi punto de vista, tampoco deberían influir en la evaluación del riesgo hechos incompletos o sesgados, como por ejemplo incomparecencias ante el tribunal pasadas que hubieran tenido alguna justificación posterior (enfermedad, imposibilidad de acudir por razones fuerza mayor que no implican falta de voluntad, etc), ya que en caso contrario, la precisión de los resultados podría verse alterada.

En España, afortunadamente, nuestra LECRim especifica y acota claramente, *stricto sensu*, en su artículo 503, qué criterios que deben tener en cuenta fiscales y jueces de instrucción para acordar medidas cautelares, en espcial, relativas a la situación personal de las personas investigadas. Y es que el juez de instrucción tiene que seguir de forma precisa la guía que la ley le proporciona para determinar la existencia de aquellos riesgos que se pretenden evitar o minimizar con la prisión provisional, en caso de querer acordarla.

Así, por un lado, para valorar la existencia de riesgo de fuga el juez deberá atender, tal y como dispone el mencionado precepto, única y conjuntamente a la naturaleza del hecho, a la gravedad de la pena que pudiera imponerse, a la situación familiar, laboral y económica de la persona investigada, así como a la inminencia de la celebración del juicio oral,

resultando posible acordar directamente la medida de prisión provisional cuando, hubieran sido dictadas al menos dos requisitorias de llamamiento y busca por cualquier órgano judicial en los dos años anteriores. No obstante, respecto de este último extremo resulta imprescindible hablar con el investigado para detectar si existen motivos que justifiquen tales incomparecencias y desvirtúen la finalidad de la norma, cuestión que como ya se ha dicho con anterioridad, debería poder tener en cuenta también, en su caso, "la máquina".

Por otro lado, para valorar el riesgo de ocultación, alteración o destrucción de pruebas, el juez deberá atender únicamente a la capacidad de la persona investigada para acceder por sí o a través de terceros a las fuentes de prueba o para influir sobre otros investigados o encausados, testigos o peritos o quienes pudieran serlo.

Y, finalmente, para valorar el riesgo de reincidencia, el juez únicamente deberá atender a las circunstancias del hecho (que deberá ser doloso), así como a la gravedad de los delitos que se pudieran cometer, dejando la posibilidad abierta, no obstante, de acordar la medida de prisión provisional sin tener en cuenta los límites penológicos legalmente previstos cuando de los antecedentes del investigado y demás datos o circunstancias que aporte la Policía Judicial o resulten de las actuaciones, pueda el juez de instrucción racionalmente inferir que este viene actuando concertadamente con otra/s persona/s de forma organizada para la comisión de hechos delictivos o realiza sus actividades delictivas con habitualidad.

En virtud de lo expuesto, pues, resulta evidente que el legislador nacional quiso acotar muy bien los factores y las circunstancias que el órgano instructor debe tener en cuenta para valorar la posible existencia de los riesgos que entraña una persona si es dejada en libertad (lo cual no solo resulta útil sino que además es garantía de seguridad jurídica).

En segundo lugar, procede analizar cómo el uso de dichas herramientas de IA en los términos previstos en la sentencia del Tribunal Supremo de Wisconsin podría chocar frontalmente con la necesidad de transparencia y expicabilidad que los derechos previstos en el artículo 24 de la Constitución Española exigen, máxime habida cuenta de que estamos ante herramientas que pueden emplearse para apoyar decisiones judiciales tan trascendentes como la privación de libertad de un investigado o acusado.
Y es que, el artículo 24 de la Constitución Española dispone:

"*1. Todas las personas tienen derecho a obtener la tutela efectiva de los jueces y tribunales en el ejercicio de sus derechos e intereses legítimos, sin que, en ningún caso, pueda producirse indefensión.*

2. Asimismo, todos tienen derecho al Juez ordinario predeterminado por la ley, a la defensa y a la asistencia de letrado, a ser informados de la acusación formulada contra ellos, a un proceso público sin dilaciones indebidas y con todas las garantías, a utilizar los medios de prueba pertinentes para su defensa, a no declarar contra sí mismos, a no confesarse culpables y a la presunción de inocencia. (…)"

Procede poner de relieve, en relación con ello, que los derechos previstos en dicho precepto, al estar incluidos en la Sección 1ª del Capítulo 2ª del Título 1º de nuestra Constitución, gozan de la calificación de fundamentales *stricto sensu* y están dotados de la especial protección (a través del recurso de amparo) conferida por el artículo 53.2 de dicha Carta Magna, que establece:

"*2. Cualquier ciudadano podrá recabar la tutela de las libertades y derechos reconocidos en el artículo 14 y la Sección primera del Capítulo segundo ante los Tribunales ordinarios por un procedimiento basado en los principios de preferencia y sumariedad y, en su caso, a través del recurso de amparo ante el Tribunal Constitucional. Este último recurso será aplicable a la objeción de conciencia reconocida en el artículo 30*."

Respecto de los mencionados derechos del citado artículo 24 de la Constitución, procede analizar cómo podría interferir en cada uno de ellos la introducción de sistemas de IA de evaluación de riesgos en nuestro proceso penal en los términos anteriormente mencionados.

Así, en primer lugar, procede poner de manifiesto que el derecho a la tutela efectiva de los jueces y tribunales en el ejercicio de los derechos e intereses legítimos de los ciudadanos es un derecho fundamental autónomo, con contenido propio y de configuración legal (tal y como se desprende de las SSTC 89/1985, de 6 de junio y 99/1985, de 20 de junio). En virtud de lo expuesto por los profesores de Derecho Jorge De Esteban y Pedro González-Trevijano, el derecho a la tutela judicial efectiva se compone, por un lado, del derecho de

libre acceso a los juzgados y tribunales o acceso a la jurisdicción, que debe incluir la posibilidad del ciudadano de dirigirse al órgano judicial competente, que debería admitir a trámite la pretensión ejercitada (con independencia de que posteriormente prospere o no), y la seguridad de que el coste del proceso judicial no va a impedir su acceso al mismo; por otro lado, del derecho a obtener una sentencia que ponga fin al litigio suscitado en la instancia adecuada; y, finalmente, del derecho al cumplimiento de la sentencia y del derecho a la interposición de los recursos legales pertinentes.[537]

Configurado tal derecho en dichos términos, se entiende que difícilmente se vería afectado por el uso de herramientas de evaluación de riesgos a través de la IA en el proceso penal español, habida cuenta de que el empleo de las mismas no tendría, *a priori*, por qué interferir en su contenido.

En segundo lugar, respecto de la cláusula de cierre relativa a la prohibición de indefensión, el Tribunal Constitucional, en su STC 48/1984, de 4 de abril establece: "*la idea de indefensión engloba, entendida en un sentido amplio, a todas las demás violaciones de derechos constitucionales que puedan colocarse en el marco del artículo 24 CE*". Y por su parte, en la STC 40/2002, de 14 de febrero, se concreta que el mencionado Tribunal "*viene declarando reiteradamente que, en el contexto del artículo 24.1 CE, la indefensión es una noción material que se caracteriza por suponer una privación o minoración sustancial del derecho de defensa; un menoscabo sensible de los principios de contradicción y de igualdad de las partes que impide o dificulta gravemente a una de ellas la posibilidad de alegar y acreditar en el proceso su propio derecho, o de replicar dialécticamente la posición contraria en igualdad de condiciones con las demás partes procesales. Por otro lado, para que la indefensión alcance la dimensión constitucional que le atribuye el artículo 24 CE se requiere [...], que la indefensión sea causada por la incorrecta actuación del órgano jurisdiccional*".

Configurado tal derecho en dichos términos, está claro que este derecho fundamental resulta vulnerable al uso de herramientas de evaluación de riesgos a través de la IA, habida cuenta de que la ausencia de transparencia y explicabilidad de las mismas o, en su caso, su

[537] Ortega, 2003.

uso extralimitado por parte de jueces y magistrados, implicaría una clara merma o privación del derecho de defensa, con un menoscabo ostensible de los principios de contradicción y de igualdad de armas procesales que impidiría o dificultaría gravemente a la defensa la posibilidad de luchar por sus intereses y rebatir la posición contraria en igualdad de condiciones.

Por su parte, en tercer lugar, en relación al derecho al juez natural, esto es, el juez ordinario predeterminado por la ley, "*exige, en primer término, que el órgano judicial haya sido creado previamente por la norma jurídica, que ésta le haya investido de jurisdicción y competencia con anterioridad al hecho motivador de la actuación o proceso judicial y que su régimen orgánico y procesal no permita calificarle de órgano especial o excepcional*" (por todas, SSTC 32/2004, de 8 de marzo; 60/2008, de 26 de mayo; y 177/2014, de 3 de noviembre).

Configurado tal derecho en dichos términos, este es uno de los derechos que más vulnerabilidad puede presentar en caso del uso de herramientas que emplean IA para evaluar riegos en el proceso penal, puesto que si la utilización de tales sistemas fuera llevada al extremo, podría acabar implicando la sustitución del juez ordinario predeterminado por la ley por una mera "máquina". En virtud de ello, en todo caso su uso tendría que ser accesorio y no debería, jamás (al menos mientras siga vigente el actual texto de la Constitución Española y la legislación que la desarrolla), sustituir el trabajo intelectual y los razonamientos del juez de instrucción, puesto que en caso contrario, como se ha dicho, este quedaría, ni más ni menos, reemplazado por un algoritmo, lo cual resultaría inconstitucional. Ello, no obstante, lo tiene también claro Tribunal Supremo de Wisconsin, que a pesar de haberse mostrado favorable al uso de tales sistemas (incluso a pesar de su opacidad), entiende, tal y como se ha avanzado con anterioridad, que la fundamentación jurídica de las resoluciones judiciales debe contener otros razonamientos que, de forma independiente y adicional al resultado del algoritmo, hayan conducido al fallo, para asegurar así que los jueces y magistrados no se basan en exclusiva en las predicciones efectuadas por los mencionados sistemas.

En cuarto lugar, respecto del derecho a la defensa (y asistencia de letrado), este es definido en la Instrucción 8/2004, de 17 de diciembre, de la Fiscalía General del Estado, como "*un*

derecho sagrado, quizás el más sagrado de todos los derechos en la justicia penal", y puede decirse que este comprende:

"*a) En primer lugar, la posibilidad de formular alegaciones en defensa de los intereses que se hayan articulado.*

b) En segundo lugar, que se garantice el derecho de probar tales alegaciones, pues es esencial en el proceso que las peticiones y alegaciones realizadas queden debidamente probadas, ya que en otro caso carecerían de valor. (…)

c) Asimismo, la garantía de la defensa debe asegurar el derecho de contradicción, ya que la misma, no es más que el derecho a la defensa que simultáneamente ejercitan las partes contrapuestas.

d) Finalmente, debemos señalar que se trata de una garantía esencial de la defensa que las pruebas y alegaciones sean tomadas en cuenta por el Juzgador, de forma que guarden congruencia entre lo pedido y alegado por las partes, y, además, que la resolución cuente con una motivación suficiente.

La defensa como garantía constitucional, asegura así, que todos los ciudadanos, cuyos intereses puedan verse afectados por una resolución judicial, tengan la posibilidad de intervenir a lo largo del proceso en el que se dicte, realizando las alegaciones oportunas y proponiendo los medios de prueba pertinentes, teniendo la posibilidad de contradecir lo propuesto por los contrarios, y dirigido todo ello a que en la resolución que se adopte, sean tenidas en cuenta y valoradas todas estas actuaciones."[538]

Y, en sentido jurídico-técnico estricto, el derecho a la defensa, tal y como se desprende de las SSTC 37/1988, de 3 de marzo, y 29/1996, de 6 de febrero, entre otras, implica la posibilidad del ciudadano de defenderse mediante la asistencia de un/a Letrado/a de su elección o, en los casos previstos legalmente, la asistencia de un/a Letrado/a del turno de oficio; y a defenderse a sí mismo.

[538] Flores & Romero, 2020, págs. 3-4.

Configurado tal derecho en dichos términos, este es, sin duda, otro de los derechos fundamentales previstos en el artículo 24 de la Constitución que más afectado podría verse por la (mala) utilización de los sistemas de evaluación de riesgo a través de la IA en el ámbito del proceso penal. Y es que, en caso de que se llevara a cabo un uso opaco e injustificado o excesivo, de tales herramientas, por un lado, tal y como ya se ha expuesto, limitaría al máximo la posibilidad de formular y probar alegaciones en defensa de los intereses articulados, puesto que tales alegaciones quedarían vacías o cojas de contenido, al desconocer la defensa las verdaderas cuestiones a debatir; y, por otro lado, el derecho de contradicción quedaría asimismo desierto de fondo, reduciéndose a un mero formalismo.

En quinto lugar, respecto del derecho del ciudadano a ser informado de la acusación formulada contra el mismo (hechos que se imputan y la calificación jurídica de los mismos), tal y como se desprende de la STC 44/1983, de 24 de mayo, entre otras, procede poner de manifiesto que resulta imprescindible para que el acusado pueda hacer uso de su derecho a la defensa. Y es que, en virtud de lo dispuesto en la STC 53/1987, de 7 de mayo, tal derecho va estrechamente relacionado con el derecho a la defensa y con el principio acusatorio que rige en nuestro ordenamiento jurídico penal: "*El principio acusatorio admite y presupone el derecho de defensa del imputado y, consecuentemente, la posibilidad de "contestación" o rechazo de la acusación. Provoca en el proceso penal la aplicación de la contradicción, o sea, el enfrentamiento dialéctico entre las partes, y hace posible el conocer los argumentos de la otra parte, el manifestar ante el Juez los propios, el indicar los elementos fácticos y jurídicos que constituyen su base, y el ejercitar una actividad plena en el proceso*".

Configurado tal derecho en dichos términos, está claro que este devendría uno de los concretos derechos fundamentales previstos en el artículo 24 de la Constitución (que, no obstante, forma parte del "todo" configurado por el derecho a la defensa y a la prohibición de indefensión) más golpeados por el uso indebido (al menos según lo previsto en nuestro ordenamiento jurídico, que desde luego no es lo mismo que lo dispuesto en el de otras jurisdicciones), de las herramientas de evaluación de riesgos a través de la IA en el ámbito penal. Y es que, justamente, la falta de transparencia y explicabilidad de tales sistemas o una extralimitación en su uso implicaría inevitablemente la imposibilidad de informar al

acusado de los verdaderos argumentos que llevan a imputarle ciertos hechos, y la defensa no podría conocer ni, por ende, rebatir la acusación.

En sexto lugar, respecto del derecho a un proceso público, sin dilaciones indebidas y con todas las garantías, procede poner de manifiesto que, por un lado, la publicidad (que debe ponerse en relación con lo dispuesto en el artículo 120.1 CE) se erige como una garantía para el ciudadano, ya que el control público tiende a evitar actuaciones dudosas o al margen de la ley; y, por otro lado, la ausencia de dilaciones indebidas (concepto jurídico indeterminado), debe evaluarse según los criterios fijados por la jurisprudencia constitucional (entre otras, SSTC 36/1984, 69/1993, 10/1997, 220/2004 y 5/2010), a saber: las circunstancias del proceso, la complejidad objetiva del mismo, la duración de otros procesos similares, la actitud procesal del recurrente, el interés que este tiene en el litigio, la actitud de los órganos judiciales, y los medios de que disponen éstos.[539]

Y en relación con el derecho a un proceso con todas las garantías, procede poner de relieve que este resulta asimismo un concepto jurídico indeterminado que ha ido siendo dotado de contenido por el Tribunal Constitucional, mediante la inclusión de múltiples garantías. En relación con ello, "*la doctrina mayoritaria y el propio Tribunal Constitucional, con una jurisprudencia en estos momentos ya consolidada, han coincidido en la respuesta última: el derecho a un proceso con todas las garantías intenta salvaguardar la presencia de ciertas instituciones específicas en la ordenación y tramitación de la realidad procesal, de tal forma que su ausencia, la falta de cualquiera de ellas, bien en la configuración legal del modo en quedebe realizarse el derecho objetivo por los órganos jurisdiccionales, bien en la propia actuación de estos últimos, origina la vulneración del derecho a un proceso con todas las garantías*."[540]

Configurados los antedichos derechos en tales términos, procede poner de manifiesto que estos podrían asimismo verse conculcados por el uso de sistemas de evaluación de riesgo a través de la IA en el ámbito del proceso penal. Y es que, por un lado, en relación al derecho a un proceso público, en caso de que los algoritmos empleados por los jueces y magistrados fueran opacos e inexplicables, o su uso fuera excesivo, la garantía de

[539] Ortega, 2003.
[540] Calderón, 2011, pág. 158.

publicidad constitucionalmente consagrada en ningún caso resultaría completa, puesto que los ciudadanos nos convertiríamos en meros espectadores de algo que, en realidad, contaría con una parte reservada, de imposible acceso. Por otro lado, respecto del derecho a un proceso con todas las garantías, hay que decir que también se podría ver sin duda frontalmente conculcado, habida cuenta de que la ausencia de transparencia y explicabilidad de los algoritmos empleados por jueces y magistrados en la toma de sus decisiones va en contra de todos los principios y garantías que constitucional y legalmente se prevén para el proceso penal.

No obstante, es de justicia poner de relieve que, en cualquier caso, el uso de sistemas de evaluación de riesgos a través de IA en el proceso judicial podría aliviar mucha carga de trabajo de los jueces y magistrados y ayudar a respetar, de una vez por todas, el tan olvidado (por desgracia y casi por obligación) derecho fundamental a un proceso sin dilaciones indebidas. Sin embargo, ello no resulta deseable "a cualquier precio", como es lógico, ya que solo un uso regulado, transparente y limitado de tales tecnologías podría devenir rentable en el ámbito de la justicia penal, ya que en caso contrario lo único que se conseguiría sería fomentar un derecho a cambio de vulnerar otro (u otros, mejor dicho) de igual o mayor entidad.

En séptimo lugar, respecto del derecho a utilizar los medios de prueba pertinentes para la defensa, la STS (Sala 2ª) 281/2009, de 18 de marzo, dispone:

"*Es obvio que el doble abordaje del derecho a la prueba -como derecho fundamental o como indebida denegación de la prueba- no altera su esencia: la quiebra se produce cuando la denegada es prueba necesaria, y por tanto es causa de indefensión en los términos del art. 24.1º de la Constitución Española (CE). Por ello es doctrina del Tribunal Constitucional que el derecho a la prueba está delimitado por cuatro consideraciones:*

a) Que la prueba sea pertinente, pues sólo a ella se refiere el artículo 24.2 CE.
b) Que dada su configuración legal, es preciso que la parte la haya propuesto de acuerdo con las previsiones de la ley procesal, es decir en tiempo oportuno y de forma legal.
c) Desde la perspectiva del Tribunal sentenciador, que éste la haya desestimado.

d) Al tratarse el derecho a la prueba de un derecho medial/procedimental que se acredite que tal denegación ha podido tener una influencia en el fallo de la sentencia, porque podría haberse variado, y es esta aptitud de la prueba denegada en relación al fondo del asunto, lo que da lugar a la indefensión que proscribe la Constitución, indefensión que debe ser material y no simplemente formal.

El derecho a la prueba no es un derecho absoluto o incondicionado, y no se produce vulneración del derecho constitucional cuando la prueba rechazada, aún siendo pertinente, carece su contenido de la capacidad para alterar el resultado de la resolución final (…)."

Configurado tal derecho en dichos términos, pues, procede poner de manifiesto que resulta evidente que quedaría conculcado en caso de que los sistemas de evaluación de riesgos empleados no resultaran transparentes y explicables, habida cuenta de que ello limitaría, *ab initio*, las posibilidades de la defensa de definir, con éxito, su completa estrategia probatoria, puesto que le resultaría imposible elegir y presentar aquellos medios de prueba que pudieran resultar más útiles y pertinentes para rebatir los motivos que llevaron a "la máquina" a fijar un determinado nivel de riesgo, lo cual no haría más que generar, una vez más, indefensión.

En octavo lugar, respecto del derecho a no declarar contra uno mismo y a no confesarse culpable, la doctrina del Tribunal Constitucional se ha mantenido unánime desde sus inicios y ha destacado tres aspectos de su contenido: "*no obligatoriedad en la declaración del presunto culpable; validez plena de las confesiones hechas de modo voluntario aunque sean contra sí mismo; no desvirtuación o invalidez de la declaración realizada en la etapa sumarial por el hecho de su rectificación en el acto del juicio oral; insuficiencia para desvirtuar el derecho a la presunción de inocencia (art. 24.2 C.E.) de la declaración contra sí mismo efectuada sin presencia de un órgano jurisdiccional*."[541]

Así, configurado tal derecho en dichos términos, entiendo que la utilización de los sistemas de IA analizados en el proceso penal podría resultar contrario al mismo, especialmente en aquellos casos en que estos se nutren, en parte, de información obtenida a través de

[541] Picó, 2012, pág. 187.

formularios que deben ser cumplimentados por la persona analizada (como por ejemplo, COMPAS), puesto que, al menos en España, el relleno de estos debería, sí o sí, resultar voluntario para ser legal (y constitucional), habida cuenta de la gran cantidad de información privada relativa al ámbito personal y social del individuo que se recopila.

En relación con ello, interesante sería debatir qué ocurriría en aquellos casos en que la persona investigada hubiera decidido rellenar el mencionado formulario en fase de instrucción y, posteriormente, decidiera acogerse a su derecho a no declarar en el plenario. ¿Qué valor podría darse, en tal caso, a los resultados del sistema de IA evaluación de riesgos que hubiera empleado los datos de dicho cuestionario para hacer sus predicciones? En mi opinión, básicamente, si la cumplimentación de dicho formulario se realizara en sede policial o bajo su mando, aun con asistencia letrada, no podría arrojar más valor que el de las declaraciones policiales; y, sin embargo, si su relleno se realizara ante el juez de instrucción (o Letrado/a de la Administración de Justicia, si la ley así lo dispusiera), con todas las garantías, podría otorgarse valor probatorio, con todas las consecuencias que ello puede conllevar en fase plenaria, como ocurriría, por ejemplo, con un informe pericial psicológico realizado en fase de instrucción con participación voluntaria del interesado (que, por supuesto, debería ser sometido a contradicción).

Y finalmente, respecto del derecho a la presunción de inocencia, la STS (Sala 2ª) 262/2017, de 7 de abril, dispone (con subrayado propio):

"*En efecto en cuanto a la vulneración del principio de presunción de inocencia en relación a la tesis defensiva del acusado, por existir alternativas plausibles razonables, en STS 681/2010 de 15 julio, con cita en las SSTS 99/2008 de 10 diciembre, 690/2009 de 25 junio, 784/2009 de 14 julio, 539/2010 de 8 junio, tenemos dicho que para determinar si esta garantía ha sido desconocida, lo que ha de constatarse en primer lugar son las condiciones en que se ha obtenido el convencimiento que condujo a la condena. Esto exige que se examine si la aportación de los elementos de la discusión sobre la aceptabilidad de la imputación se efectúa desde el respeto al método legalmente impuesto, de suerte que los medios de prueba sean considerados válidos y el debate se someta a las condiciones de contradicción y publicidad.*

En segundo lugar, como también indicábamos en aquellas resoluciones, y como contenido específico de este derecho fundamental a la presunción de inocencia, deberá examinarse si, prescindiendo del grado de seguridad que el Juez tenga sobre el acierto de su convicción, ese método ha llevado a una certeza objetiva sobre la hipótesis de la acusación. No porque se demuestre una verdad indiscutible de las afirmaciones que funda la imputación. Sino porque, desde la coherencia lógica, se justifique esa conclusión, partiendo de proposiciones tenidas indiscutidamente por correctas. Para constatar el cumplimiento de este específico presupuesto de enervación de la presunción constitucionalmente garantizada han de verificarse dos exclusiones:

La primera que la sentencia condenatoria no parte del vacío probatorio, o ausencia de medios de prueba, que aporten proposiciones de contenido incriminador y sean válidamente obtenidas y producidas en el debate oral y público. El vacío habrá sido colmado cuando, más allá del convencimiento subjetivo que el Juez, al valorar los medios de prueba, adquiera sobre la veracidad de la acusación, pueda estimarse, en trance de revisión, que no sustitución, de la valoración del Juez, que los medios que valoró autorizan a tener por objetivamente aceptable la veracidad de la acusación o, si se quiere, a excluir la mendacidad de la acusación.

La segunda la inexistencia de alternativas, a la hipótesis que justificó la condena, susceptibles de ser calificadas como razonables. Y ello porque, para establecer la satisfacción del canon de razonabilidad de la imputación, además, se requiere que las objeciones oponibles se muestren ya carentes de motivos racionales que las justifiquen de modo tal que pueda decirse que excluye, para la generalidad, dudas que puedan considerarse razonables. Ahora bien ello no implica que el Tribunal esté obligado a considerar probadas todas las alegaciones formuladas por el acusado, ni que tenga que realizar un análisis exhaustivo de cada una de las pruebas practicadas, lo que sí está obligado que es a ponderar y valorar la prueba de descargo junto con la de cargo, lo que representa un presupuesto sine qua non indispensable para que el juicio de autoría pueda formularse con la apoyatura requerida por nuestro sistema constitucional. No se trata, claro es, de abordar todas y cada una de las afirmaciones de descargo ofrecidas por la parte pasiva del proceso (STS 258/2010 de 12 marzo , 540/2010 y 8 junio). En palabras del Tribunal Constitucional exige solamente ponderar los distintos elementos probatorios,

pero sin que ello implique que esa ponderación se realice de modo pormenorizado, ni que la ponderación se lleve a cabo del modo pretendido por el recurrente, sino solamente que se ofrezca una explicación para su rechazo (SSTC. 187/2006 de 19 junio , 148/2009 y 15 junio)."

Configurado tal derecho en dichos términos, considero que la utilización de los analizados sistemas de IA en el proceso penal podría, sin duda, implicar su vulneración. Y es que, por un lado, es evidente que el derecho a la presunción de inocencia lleva implícita una presunción "*iuris tantum*" que solamente puede ser destruida por indicios racionales y pruebas de cargo con entidad suficiente para ello y, en mi opinión, desde luego, por sí solo el mero resultado de un programa de *software* de evaluación de riesgos no podría obtener tal calificación. Y, por otro lado, aunque la antedicha evaluación de riesgos fuera considerada junto con otras pruebas de cargo o indicios racionales para tomar según qué decisiones judiciales, entiendo que debería quedar muy claro, a través de sistemas transparentes y explicables, que los datos de los que estos se nutren tienen base legal y no resultan discriminatorios, puesto que en caso contrario podría llevarse a cabo una frontal e injusta vulneración del derecho a la presunción de inocencia de aquellos ciudadanos que cuenten con ciertas características históricamente sesgadas que el algoritmo considere determinantes de peligro.

En relación con todo lo anterior, no obstante, procede poner de manifiesto que en el caso español, bajo mi punto de vista, la ausencia de transparencia y explicación de los algoritmos empleados en el ámbito del proceso penal jamás podría quedar justificada o amparada por el secreto comercial o de empresa alegado por las compañías propietarias de los sistemas en EEUU. Y es que, dada la configuración constitucional vigente, el derecho al secreto de empresa en todo caso vencería frente a los derechos proclamados en el artículo 24 de la Constitución Española, siendo que estos segundos, como se ha dicho, son considerados por nuestro Tribunal Constitucional como derechos fundamentales *stricto sensu*, por quedar incluidos dentro de la Sección 1ª del Capítulo 2ª del Título 1º de la Carta Magna y, sin embargo, el primero (que no está previsto ni siquiera como tal pero se entiende que podría quedar incluido en el derecho a la libertad de empresa en el marco de la economía de mercado reconocido por el artículo 38 de la Carta Magna), no ostenta tal

consideración, puesto que está ubicado en la Sección 2ª del Capítulo 2ª del Título 1º y, por ende, tiene la consideración de derecho fundamental *lato sensu*.
A la vista de lo expuesto, el avance en el uso de las tecnologías de IA de evaluación de riesgos en el ámbito judicial, que en los Estados Miembros de la UE todavía están por implementar, tal y como ya se ha anunciado, de forma inevitable debe ir ligado al avance de la legislación sobre las mismas.

Una vez más, pues, nos hallamos ante la clamorosa y urgente necesidad de que se unifiquen esfuerzos y criterios y se sienten unas bases legales claras y sólidas por parte de los Estados (o, en su caso, la UE) sobre las que las empresas privadas que desarrollan herramientas de IA (que, sin duda, pueden resultar muy útiles) puedan operar, ya que, en mi opinión, la legislación actual resulta abslutamente insuficiente. Y es que bajo mi punto de vista, la previsión legal expresa es la única forma que tenemos de avanzar hacia un proceso penal garantista y seguro que pueda recibir ayuda de la tecnología para ganar eficacia, debiendo tener claro que el uso de sistemas de IA de evaluación de riesgos en el ámbito judicial con garantías legalmente previstas abriría sin duda la puerta al progreso.

En relación con ello, afortunadamente, no obstante, actualmente ya está en marcha la elaboración de un Reglamento sobre IA que con alta probabilidad regulará el uso de los sistemas analizados, tal y como se desprende de la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, tal y como ya se ha expuesto con anterioridad.

Es importante poner de relieve, no obstante, que a pesar de que entre en vigor la mencionada legislación europea, que sin duda supondrá una garantía de gran utilidad, habida cuenta del estricto control al que se pretende someter a los sistemas de IA calificados de alto riesgo, entre los que, como ya se expuso al hablar de los sistemas policiales de evaluación de riesgos[542], se encuentran las herramientas de evaluación de riesgos analizadas, los jueces y magistrados no deberán bajar la guardia a la hora de utilizarlas en su día a día. Y es que, incluso en tales circunstancias, aunque el uso de dichas

[542] Véanse págs. 213-214.

herramientas de IA devenga certificado como seguro y legítimo, su empleo por parte de los jueces para tomar decisiones deberá realizarse con la máxima cautela para proteger, principalmente, el antedicho derecho del juez ordinario predeterminado por la ley y el resto de derechos en juego, al igual que los médicos deben tener máximo rigor y cuidado a la hora de dispensar los distintos medicamentos en cada caso, por mucho que cuenten con un certificado de calidad. Me explico.

Los jueces son humanos y, como tales, tienen sus fortalezas, únicas, pero también sus debilidades, y estas pueden conllevar especiales riesgos cuando se enfrentan al uso de los sistemas algorítimicos de evaluación de riesgos. No es ningún secreto que la naturaleza, especialmente la humana, por lo general, hace tender a los seres humanos a realizar el mínimo esfuerzo para conseguir el mayor beneficio, ya que los recursos tanto mentales como físicos son limitados y, en última instancia, prima la supervivencia. Dicho esto, no es de extrañar que, si un juez tiene a su alcance herramientas que le permiten hacer un menor esfuerzo intelectual y le llevan a conseguir un buen resultado, estará como mínimo tentado a darles un uso intenso y generalizado, librándose incluso, en ocasiones, de realizar ciertos trabajos, delegándolos en "la máquina", especialmente en juzgados con grandes (y muchas veces insoportables) cargas de trabajo. Ello, no obstante, implica un elevado peligro para los derechos fundamentales de los justiciables y, en especial, para el derecho al juez ordinario predeterminado por la ley, para el derecho a la defensa y para el derecho a un proceso con todas las garantías.

Así, por un lado, si los jueces más inseguros o perezosos (y, por qué no, los menos profesionales) tienen la posibilidad de acudir directamente a programas de IA de evaluación de riesgos con carácter previo a analizar cada asunto y pueden basarse en lo que estos dispongan para tomar las decisiones, tenderán a confiar en gran medida en ellos y a minimizar su trabajo intelectual. Y es que en la práctica totalidad de ocasiones lo realmente difícil en un asunto es tomar la decisión final, ya que casi siempre suelen existir argumentos a favor y en contra de lo que uno mismo piensa. Así, en caso de que un programa de IA fuera el que decantara la balanza de modo inicial, al juez le resultaría relativamente sencillo realizar su trabajo, puesto que evitaría la parte intelectualmente más desafiante y compleja, que es la de resolver el asunto, y siempre podría plasmar en la resolución, *a posteriori*, razones traídas *ad hoc* que sustentaran la decisión (previamente tomada por "la máquina").

De tal forma, los justiciables jamás conocerían la verdad escondida tras la toma de la decisión judicial, puesto que aparentemente el algoritmo certificado no sería más que un elemento decisorio más, pero la realidad es que este habría sido el verdadero elemento decisor y que el juez únicamente se habría encargado de buscar algunos argumentos que lo sustentaran *a posteriori* (lo contrario de lo que debería ser). Y ello no supondría avance alguno, sino un claro paso hacia atrás en el respeto de los derechos de los ciudadanos y la justicia.

Por otro lado, sin embargo, los jueces más diligentes y trabajadores (y, por ende, los más profesionales) aunque tuvieran a su alcance programas de IA de evaluación de riesgos, optarían sin duda por analizar los casos con carácter previo y reflexionar y buscar las razones que les llevaran a tomar su propia decisión, y solo acudirían a la tecnología al final, a modo de confirmación, validación o comparación con lo ya decidido, es decir, no como factor determinante, si no "a mayor abundamiento".

En este último escenario, no obstante, surgen algunas dudas de difícil respuesta. ¿Qué ocurriría en caso de que la decisión previamente tomada por el/la juez/a fuera contraria a lo indicado por el algoritmo? ¿Tendría el/la juez/a la posibilidad real de rebatir o exponer sus diferencias con el resultado del programa de IA (lo cual sería, sin duda, lo deseable)?. En mi opinión, a la vista de ello, la única forma en que una herramienta de IA podría tener encaje constitucional y legal (salvo que haya cambios), sería aquella en que se respetara al máximo nivel el poder decisorio del juez, y ello únicamente sería posible en caso de que este pudiera conocer realmente el contenido del algoritmo de evaluación del riesgo (para lo cual este debería ser transparente y, sobre todo explicable) y tuviera la posibilidad de rebatir o secundar, con sus razonamientos, la decisión de "la máquina". Así, solo en caso de que el juez supiera qué datos han sido empleados por el algoritmo, qué factores han sido tenidos en cuenta por este, qué peso se ha otorgado a cada uno de ellos en la decisión, etc, podría, a través de su necesaria argumentación, motivar el porqué de su apoyo o discrepancia con la decisión del sistema y detectar, en su caso, posibles puntos constitucional o legalmente dudosos. Y es que considero importantísimo, a los efectos de que siempre conste cuál ha sido el verdadero trabajo intelectual del juez, que este plasme en sus resoluciones aquellas razones que le hayan llevado tanto a estar de acuerdo con el

resultado algorítmico, en caso de secundarlo, como en desacuerdo, en caso de rechazarlo, siendo este valorado como una prueba más.

No cabe duda, no obstante, de que en determinadas ocasiones, la discepancia entre la valoración realizada por el juez y por "la máquina" llevaría a este a replantearse su decisión, ya que es inevitable que exista una cierta influencia mental al ver que lo que resulta de analizar millones de datos de forma automática (infinitos más de los que la mente humana puede llegar a procesar) es distinto de lo se desprende de su particular y limitado análisis mental. Ello seguro que llevaría al juez, al menos, a revisar su decisión, si bien en caso de que el contenido del algoritmo fuera transparente y explicable, el debate que surgiría sería legítimo y "sano", en cualquier caso. Y es que el juez, tal y como se ha venido advirtiendo, es humano y, por ende, inevitablemente influenciable y, al igual que en el caso de que tuviera acceso a una prueba que posteriormente fuera declarada ilícita podría tener dificultades para obviar mentalmente su contenido, en el supuesto de ver el resultado de un algoritmo que analiza en segundos millones de datos a los que él no tiene acceso, también podría verse inmerso en un complejo ejerecicio de disociación.

Para mejor ilustración, resulta interesante poner un ejemplo práctico (llevado al extremo), que permita comprender mejor a qué se está haciendo referencia.

Así, imaginemos que llega al Juzgado de Guardia un detenido que ha cometido un robo con fuerza en el interior de un vehículo, que cuenta con un amplio historial delictivo, especialmente de delitos violentos, y que ha sido arrestado *in fraganti* por la policía, razones por las que el Ministerio Fiscal solicita la celebración de la comparecencia de prisión provisional prevista en el artículo 505 de la LECrim. Imaginemos, también, que el detenido no cuenta a su Letrado que tiene problemas de adicción a sustancias estupefacientes y que, por ende, este no puede trasladar tal información al juez.

En el primero de los casos expuestos, llevado al extremo, repito, el juez de guardia, con ánimo de realizar un mínimo esfuerzo y ventilar el asunto con la mayor agilidad, al recibir el atestado -que pondría de manifiesto que el detenido fue arrestado en el momento en que se disponía a huir del vehículo con el material sustraído, que estaba muy agresivo y que se resistió a la detención-, lo leería e inmediatamente después acudiría al programa de IA para

que le diera una rápida predicción del riesgo de reiteración delictiva del detenido, que en este caso, casi seguro, resultaría elevado. Así, dicho juez (habida cuenta de la detención *in fraganti*) con toda probabilidad simplemente acordaría tomar declaración -sin profundizar en exceso- al detenido, que posiblemente negaría los hechos o se acogería a su derecho a no declarar, asesorado por su Letrado; pospondría la práctica de más diligencias de instrucción para un momento ulterior del procedimiento (entre otras, por ejemplo, la de escuchar como testigo a la mujer de este y a una vecina que vio lo ocurrido desde el balcón, que no constarían citadas para ese día, cuyos datos constarían en el atestado), puesto que ya tendría indicios suficientes para decidir sobre su situación personal; y celebraría la solicitada comparecencia de prisión, tras lo cual dictaría un Auto de prisión provisional alegando riesgo de reiteración delictiva por los múltiples antecedentes penales del detenido, por la flagrancia de los hechos y la ausencia de justificación o versión exculpatoria.

En el segundo de los casos expuestos, no obstante, el juez de guardia, al recibir el atestado, lo leería e inmediatamente después pensaría qué diligencias de investigación sería pertinente acordar y practicar no solo para instruir el caso sino también para decidir sobre la situación personal del detenido. Tras ello, imaginemos que el juez acordara ese mismo día tomar declaración testifical a la mencionada esposa del investigado y esta le explicara que su marido padecía un problema grave de adicción a sustancias estupefacientes desde hace tiempo, que estaba siguiendo un programa de rehabilitación que aperentemente le había dado buenos resultados hasta la semana anterior, en que falleció su madre y "volvió a las andadas", estando completamente fuera de sí el día de los hechos. Imaginemos también que esta se mostrara absolutamente dispuesta a colaborar y aportara al juez los datos del centro de desintoxicación que trataba a su marido para intentar que este acordara un ingreso involuntario del mismo o articulara algún tipo de cooperación. Imaginemos, asimismo, que se acordara la declaración testifical de la vecina para ese mismo día y que, de forma compatible con lo anterior, explicara al juez que un par de horas antes de los hechos había escuchado cómo el detenido hablaba por teléfono desde la calle, en evidente estado de ansiedad y falta de autocontrol, y decía a su interlocutor que iba a conseguir a toda costa algo de "género" para darle a cambio de "esas rayitas de cocaína que tanto necesito, porque si no, literalmente me muero". Tras ello, imaginemos que el juez, con una perspectiva sobre lo ocurrido absolutamente distinta a la del caso anterior, tomara

declaración al detenido y le preguntara por los hechos expuestos, este se derrumbara y finalmente los reconociera, solicitándole incluso ayuda para poder volver al centro de rehabilitación donde estaba siguiendo tratamiento, admitiendo que la muerte de su madre le había vuelto a llevar por el mal camino, de forma puntual, mostrándose completamente dispuesto a colaborar. En tal caso, seguramente, el juez, a pesar de que la herramienta de evaluación de riesgos arrojara un resultado preocupante, remitiría un oficio o contactaría de algún modo con el centro de rehabilitación mencionado y, con la ayuda de la esposa, podría pactar un reingreso voluntario del detenido, seguramente con el beneplácito del Ministerio Fiscal, que podría llegar a retirar su petición de prisión provisional al entender que la medida de internamiento (en este caso, voluntario) en un centro de desintoxicación podría ser la más adecuada y podría, además, alcanzar un pacto con la defensa para que se dictara una Sentencia de conformidad, con suspensión de la pena de prisión, en virtud de lo dispuesto en el artículo 80.5 del Código Penal.

Y, *voilà*, así es como queda claramente patente la necesidad de que los jueces no deleguen en "las máquinas" las decisiones que afecten a los justiciables, especialmente en un ámbito tan sensible como es el de la evaluación de riesgos para conceder o no la libertad, puesto que en muchas ocasiones los casos no están compuestos únicamente por "datos", sino que hay una parte más profunda y humana, que va mucho más allá y que solo sale a la luz haciendo un esfuerzo intelectual más intenso y puede tener un peso fundamental en su resolución.

En el supuesto expuesto, en concreto, además, no está claro que el juez del primer caso hubiera acabado llamando a declarar como testigos, en fase de instrucción, a la vecina y a la esposa del investigado, puesto que este ya había sido detenido "*in fraganti*", sin que existieran dudas sobre su autoría y, por ende, el instructor podría considerar que disponía de indicios suficientes para seguir adelante, dictar Auto de continuación de Procedimiento Abreviado y acordar, incluso, la apertura del juicio oral. Tampoco está claro, además, si en el acto del juicio oral se hubieran propuesto tales declaraciones testificales por el Ministerio Fiscal (y, en su caso, por la acusación particular) o la defensa, o incluso si hubieran sido admitidas, lo que hubiera llevado, casi sin duda, a una condena del acusado por delito de robo con fuerza, con la consiguiente pena de privación de libertad. Y es que, incluso en caso de que el juez instructor hubiera acabado escuchando en declaración a la esposa y/o

la vecina del investigado en fase de instrucción, con carácter posterior a haber dictado Auto de prisión provisional, el daño (menor, por supuesto) ya se hubiera producido, habida cuenta de que, a lo mejor, el mencionado autor de los hechos simplemente necesitaba una oportunidad para reconducir su camino de recuperación (la opción más beneficiosa para él y para toda la sociedad, sin duda), interrumpido de forma puntual por una circunstancia excepcional (la muerte de su madre) y, sin embargo, el ingreso en prisión hubiera facilitado su recaída y dificultado su ulterior rehabilitación.

En relación con lo expuesto, una vez más procede advertir que para que todas las partes de un procedimiento tengan conocimiento real de cuáles han sido las verdaderas razones que han llevado a un juez a tomar una determinada decisión, tengan garantía de que este no ha sido sustituido por una "máquina", y puedan tener la posibilidad real de rebatir sus argumentos, es absolutamente imprescindible y fundamental que exista la mayor transparencia posible en las actuaciones judiciales. No obstante, ello no resulta una tarea fácil, ya que tal y como se ha advertido, incluso en aquellos casos en que un juez delegue, *de facto*, en un programa de *software* la resolución de un asunto, este puede buscar argumentos *ex post* que "vistan" y justifiquen su decisión de forma aparentemente humana y legítima. Y es que, tal y como puso de manifiesto Frank Pasquale, Profesor de Derecho de la Universidad de Maryland (EEUU), de forma bastante acertada, en mi opinión (especialmente en relación a los jueces estadounidenses), los jueces tienen la opción de rechazar la conclusión del algoritmo y pueden seguirla únicamente en aquellos casos en que les sirva como una "excusa" conveniente.[543] En cualquier caso, la polémica está servida.

No obstante, podría decirse que, en un ámbito ajeno a la IA, una problemática similar ya se da en la actualidad y se ha venido dando a lo largo de todos los tiempos, desde que los jueces existen. Y es que, al igual que los sistemas de IA, los jueces humanos también tienen su particular *black box*, que no es otro que su cerebro. Así, si bien un juez puede armar de forma jurídicamente impecable una resolución y exponer de forma clara y ordenada los motivos que le han llevado a tomar la correspondiente decisión, lo cierto es que en muchas ocasiones esos argumentos plasmados en el papel en realidad esconden razones (creencias

[543] Van Dam, 2019.

religiosas, convicciones personales o sesgos, entre otros) que jamás trascenderán al público y que quedarán ocultas en la mente del magistrado. El problema, en tal caso, es mayor incluso que en el de los algoritmos, habida cuenta de que, con la tecnología actual, es perfectamente viable exigir que estos devengan transparentes y explicables para poder ser empleados con garantías en el ámbito de la justicia, si bien ello no puede ser objetivamente requerido a los jueces, ya que el contenido de sus pensamientos (todavía) no puede ser objeto de conocimieno y escrutinio público. Así, si un juez, a la hora de tomar una decisión, se ve influenciado (consciente o inconscientemente), principalmente, por el hecho de que la persona investigada sea de género masculino, de raza negra o de religión musulmana, pero luego argumenta su resolución con motivos aparentemente legítimos, nunca nadie podrá conocer ni rebatir las verdaderas razones que le llevaron a dictar tal fallo, lo cual resultaría igual de contrario al derecho a la defensa y a un proceso con todas las garantías debidas que el hecho de que un algoritmo sea opaco y se combine con otros factores de decisión (incluso mucho más, puesto que ni siquiera existiría la posibilidad de alegar tal opacidad por la defensa, salvo en casos absolutamente flagrantes o extremos). Ello, sin embargo, suele obviarse (salvo en casos de extrema flagrancia), como consecuencia de la presunción de profesionalidad y, sobre todo, independencia, de los jueces y magistrados que integran la Carrera Judicial, que han prestado juramento o promesa de cumplir y hacer cumplir la Constitución y las leyes.

En el momento actual, no obstante, es importante poner de manifiesto que, lamentablemente, la anhelada transparencia de los sistemas de evaluación de riesgos puede que no siempre resulte posible y efectiva, ya que tal y como afirma Nyssa Taylor, asesora de políticas de justicia penal de la organización ACLU[544] (EEUU): "*Incluso si los gobiernos comparten cómo los sistemas toman sus decisiones -lo que ya sucede en Filadelfia en algunos casos-, las matemáticas a veces son demasiado complejas para que la mayoría de las personas lo entiendan*."[545] No obstante, en mi opinión, con toda probabilidad, estoy segura de que a medida que avance el uso certificado de tales sistemas de IA su explicabilidad irá convirtiéndose en un requisito legal y, por ende, aunque pueda resultar muy complicada para los jueces, fiscales, letrados y demás ciudadanos su

[544] American Civil Liberties Union of Pennsylvania, s.f..
[545] Metz & Satariano, 2020.

comprensión, existirán peritos que se dediquen a "traducirles" aquello que hay detrás de las decisiones algorítmicas empleadas en el proceso penal.

Dicho esto, procede poner de manifiesto que, en cualquier caso, una herramienta de evaluación de riesgos que emplea IA, siempre que resulte objetiva, transparente y explicable y cumpla con los estándares legales (previa certificación por un organismo competente e independiente), sería de enorme utilidad tanto para los ciudadanos, que verían incrementada la seguridad jurídica, habida cuenta de que se tenderían a unificar criterios, dejando atrás la mayor parte de los sesgos y errores todavía existentes (humanos) en los sistemas judiciales; como para los jueces y magistrados, habida cuenta de que dichos sistemas tienen la capacidad de evaluar millones de variables y de casos más que los que limitada y humanamente pueden examinar estos (que desde luego, para acordar una medida cautelar se basan en las circunstancias del caso concreto pero también en las máximas de su experiencia), por lo que podrían ver reforzado su trabajo y recibir asistencia para impartir una justicia de mayor calidad.

3.2.2. HERRAMIENTAS DE INVESTIGACIÓN CRIMINAL

3.2.2.1. Concepto

Las herramientas de IA de investigación criminal propiamente dichas pueden ser definidas como aquellos sistemas que emplean tal tecnología y se utilizan por las autoridades policiales, fiscales y judiciales para averiguar y hacer constar la perpetración de los delitos, con todas las circunstancias que puedan influir en su calificación, así como la culpabilidad de los delincuentes.[546]

Procede poner de manifiesto, no obstante, que si bien la práctica totalidad de las herramientas de IA que se analizarán en la presente Sección no fueron originariamente diseñadas con la finalidad de ser destinadas a la investigación criminal, ya que no existe una categoría científica tan específica en el ámbito de dicha tecnología, lo cierto es que

[546] En relación con lo dispuesto en el artículo 299 de la LECrim.

cuentan con unas características y unas utilidades prácticas que, sin duda, las hacen idóneas para conseguir tal objetivo.

Y es que parte de mi labor de investigación ha consistido, justamente, en rastrear y analizar las múltiples y distintas herramientas existentes en el ámbito de la IA y hacer una cuidada y específica selección de aquellas que he creído más adecuadas y aptas para ser puestas a disposición de las autoridades policiales, fiscales y judiciales con el fin de asistirles en la investigación de delitos.

Tal y como ya se ha anunciado en la Introducción de la presente tesis doctoral, la finalidad de ello no es otra que la de dotar a la Administración Pública de mejores y más sofisticados medios técnicos que sirvan para auxiliar y dar soporte a los ya existentes y a las personas que están al servicio de la investigación de las causas penales, ya que, al menos en España, como es bien sabido, la justicia permanece colapsada y, en muchos aspectos, estancada y obsoleta, lo que impide llevar a cabo instrucciones de calidad no solo más rápidas sino también con la máxima excelencia.

Si bien la policía judicial de nuestro país[547] está dotada de medios de investigación criminal más sofisticados que los que tienen a su alcance las autoridades fiscales y judiciales, lo cierto es que, tras las pesquisas llevadas a cabo, he llegado a la conclusión de que en muchos aspectos tecnológicos (especialmente, en materia de IA) esta se halla a la cola de Europa, lo cual, no obstante, trata de ser compensado por la enorme profesionalidad y experiencia de sus miembros, pero en cualquier caso resulta insuficiente.

3.2.2.2. Clases

Una vez establecido el concepto de lo que, a mi entender, son los sistemas de IA de investigación criminal (ya que, tal y como he advertido, no existe una categoría *per se* calificada como tal), procede entrar a analizar cada una de las distintas herramientas que, como he dicho, considero que pueden tener utilidad para tal fin y, por ende, pueden resultar englobadas en dicha clase.

[547] Que ostenta la labor de auxiliar a las autoridades fiscales y judiciales en la investigación delictiva, en virtud de lo dispuesto en el artículo 282 de la LECrim.

Con el objetivo de sistematizar y dar una mayor uniformidad al examen de cada una, no obstante, he decidido seguir un mismo orden de análisis para todas ellas, a saber: concepto, subclases, posibles utilidades en la instrucción de las causas, regulación, y potenciales riesgos jurídicos que su uso puede entrañar.

Empezamos.

A-Herramientas de IA que emplean datos biométricos

A.1. *Concepto*

Los datos biométricos son definidos en la legislación europea, en concreto por el artículo 3.13 de la Directiva sobre protección de datos en el ámbito penal, así como el artículo 4.14 del Reglamento General de Protección de Datos (RGPD) y el artículo 3.18 del Reglamento (UE) 2018/1725, que disponen que son "*datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos*"; y por la española, en concreto por el artículo 5.l) de la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, que establece que son "*datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o de conducta de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos*", siendo la biometría la ciencia que se encarga de su análisis, con fines principalmente identificativos.

Y es que el valor de los datos biométricos humanos, únicos y característicos de cada individuo, generalmente intransferibles, fácilmente reconocibles y comprobables, es inmenso, especialmente en el ámbito de la investigación criminal.

En relación con ello, es interesante poner de relieve que la biometría opera mediante dos tipos de técnicas:

a) mediciones fisiológicas, que incluyen el análisis de características físicas únicas de los seres humanos propiamente dichas, tales como la forma del rostro, el iris, las huellas dactilares o el ADN; y

b) mediciones de comportamiento, que incluyen el análisis de características conductuales únicas de los seres humanos, tales como la voz, los gestos, la firma, la escritura, etc.

Por lo general, si bien ambas técnicas de identificación biométrica cuentan con elevados grados de precisión, la primera sin duda es la que arroja unos resultados más fiables, habida cuenta de la mayor inmutabilidad que experimentan las características humanas analizadas. Y es que, mientras las mencionadas singularidades físicas suelen mantenerse inalterables durante toda la vida del ser humano, las características de comportamiento están sujetas a mayores cambios e influencias externas, tales como por ejemplo, los nervios o el estrés. No obstante, incluso dentro de la primera categoría existen diferentes niveles de precisión, habida cuenta de que el ADN de una persona, por ejemplo, es el mismo y permanece invariable durante toda su vida, lo que no sucede, sin embargo, con el rostro, que va experimentando cambios a lo largo del tiempo. Así pues, queda claro que a pesar de que la biometría es una ciencia que abarca el estudio de todos los datos biométricos de que dispone un individuo humano, esta no cuenta con un grado de infalibilidad uniforme, sino que varía en función de la información analizada.

En relación con lo expuesto, procede poner de manifiesto que, si bien la biometría no es una ciencia precisamente moderna,[548] lo cierto es que en los últimos años ha experimentado grandes avances gracias a su combinación con la IA. Y es que la aparición de sistemas de *Machine Learning* y *Deep Learnig* capaces de ser entrenados con millones de datos para efectuar análisis automáticos de información biométrica con fines identificativos y comparativos ha revolucionado tal antigua ciencia. Así, hoy en día existen herramientas de IA capaces de captar datos biométricos (a partir de una imagen o de una grabación de voz, por ejemplo), compararlos con una cantidad ingente de información biométrica contenida en vastas bases de datos, detectar patrones y coincidencias a través de un algoritmo

[548] Ya en la prehistoria, por ejemplo, el ser humano firmaba con el dedo.

entrenado para ello, e identificar o verificar la identidad de un individuo en cuestión de segundos, por ejemplo.

No obstante, la biometría combinada con la IA no es una ciencia exacta al 100%, al menos por el momento, habida cuenta de que se basa en algoritmos estadísticos que pueden dar lugar a falsas apreciaciones tanto por exceso como por defecto, tales como por ejemplo la asociación de dos rasgos biométricos que en realidad no corresponden al mismo individuo, o la ausencia de detección de algún rasgo biométrico concreto de una persona. En cualquier caso, no obstante, la precisión de los sistemas suele ir ligada con la calidad de los mismos y, especialmente, con la naturaleza de los datos que los nutren y el nivel de entrenamiento de los algoritmos encargados de procesar la información.

Con el fin de reducir al máximo las tasas de error, no obstante, cada vez son más los sistemas de IA que combinan el análisis de al menos dos tipos de datos biométricos distintos. Así, no resulta infrecuente toparse con herramientas de IA multimodales que, a diferencia de las unimodales, analizan a la vez, por ejemplo, las huellas dactilares y el rostro de las personas, como ocurre en la mayoría de las máquinas de identificación automática colocadas en los aeropuertos[549]. Y es que ello es un modo de reforzar la seguridad y la fiabilidad de los resultados que arrojan los sistemas, habida cuenta, especialmente, de la proliferación de las falsificaciones de datos biométricos y la creación de los denominados "datos biométricos sintéticos" por parte de los criminales.

En relación con esto último, procede poner de manifiesto que si bien, hasta ahora, una de las principales y más valiosas características de los datos biométricos era su singularidad y la ausencia de posibilidades de falsificación y sustracción (lo que los diferenciaba de otras formas de identificación tales como las contraseñas, por ejemplo), lo cierto es que hoy en día ello está cambiando. Y es que bien es sabido que a pesar de que la IA debe tender a usarse para facilitar y mejorar la vida de los seres humanos, en ocasiones esta es empleada con otros fines menos nobles. Y eso es lo que ha ocurrido, por ejemplo, con las denominadas *DeepMasterPrints*, unas huellas dactilares sintéticas creadas por una red neuronal de IA entrenada para ello capaces de suplantar la identidad de las personas.[550]

[549] Las denominadas "*Automated Border Control*" (ABC).
[550] Véase Bontrager, Roy, Togelius, Memon & Ross, 2018.

Dicho lo anterior, es importante poner de manifiesto que los sistemas de IA que emplean datos biométricos suelen tener una de las siguientes funciones:

-por un lado, la *identificación propiamente dicha*, que tiene como objetivo averiguar y determinar la identidad de una persona mediante el cotejo de sus datos biométricos (dubitados) con una gran cantidad de información de la misma clase (indubitada) que se halla almacenada en una base de datos (o varias vinculadas), lo que responde a la pregunta de: ¿quién es tal individuo?; y

-por otro lado, la *comprobación o verificación*, que tiene como finalidad constatar o rechazar la identidad de una persona mediante la comparación o superposición de dos o más datos o conjuntos de datos biométricos (dubitados e indubitados), lo que responde a la pregunta de: ¿es realmente este individuo quien dice ser o quien creemos que es?.

En el ámbito de la UE varios son los proyectos que se están desarrollando con el fin de emplear los datos biométricos con distintos fines, entre otros, el refuerzo de la seguridad.

En relación ello, especialmente relevante es la adopción por parte de la UE de dos Reglamentos (en vigor desde el 11 de junio de 2019) que establecen un marco jurídico para la interoperabilidad de los sistemas de información en el ámbito de la justicia y de la seguridad: el Reglamento (UE) nº2019/817 del Parlamento Europeo y del Consejo, de 20 de mayo de 2019, relativo al establecimiento de un marco para la interoperabilidad de los sistemas de información de la UE en el ámbito de las fronteras y los visados; y Reglamento (UE) nº2019/818 del Parlamento Europeo y del Consejo, de 20 de mayo de 2019, relativo al establecimiento de un marco para la interoperabilidad entre los sistemas de información de la UE en el ámbito de la cooperación policial y judicial, el asilo y la migración.

Y es que el fin de tales instrumentos legales no es otro que el de facilitar el intercambio de información entre Estados Miembros de la UE con el objetivo de, por un lado, mejorar la seguridad mediante el cruce de datos de todos aquellos que traspasan nuestras fronteras (prestando especial atención a los individuos con múltiples identidades) y, por otro lado, controlar los flujos migratorios y combatir la inmigración ilegal, siempre con plena salvaguarda de los derechos fundamentales. Así, las autoridades competentes de los

distintos Estados Miembros pueden (o podrán, en su caso) acceder de forma simultánea a los diversos sistemas de información europeos que posteriormente analizaré con más detalle (tres sistemas de información centralizados de la UE ya existentes: el SIS, el VIS y Eurodac; tres sistemas de información centralizados de la UE, en desarrollo: el SES[551], el SEIAV[552] y el ECRIS-TCN; y la base de datos de Europol) que contienen extensísimas bases de datos biográficos y biométricos (rasgos faciales y huellas dactilares, principalmente) y, asimismo, buscar, analizar y comparar estos últimos.[553] [554]

De acuerdo con lo expuesto, procede entrar a analizar más en profundidad el concepto de cada una de las distintas herramientas de IA que emplean datos biométricos, sus posibles utilidades en el ámbito del proceso de instrucción penal y sus potenciales riesgos.

A.2. *Subclases*

a) Reconocimiento facial

a.1) Concepto

A pesar de que la mayoría de la gente piensa que el reconocimiento facial es una tecnología lejana, futurista y sofisticada, lo cierto es que esta nos acompaña en muchos momentos de nuestro día a día, incluso a veces sin que tengamos conciencia de ello. Y es que tal técnica de IA basada en el entrenamiento de la "vista" de las máquinas se emplea, por ejemplo, por nuestros *smart phones* para reconocer nuestra identidad y permitirnos acceder a su contenido o pagar en el supermercado con una tarjeta de crédito previamente introducida en el sistema; por la red social Facebook para etiquetarnos en fotografías; o por la red social Instagram para detectar el lugar en que ha de colocarse el filtro que seleccionamos antes de publicar una imagen.

Conceptualmente, puede decirse que el reconocimiento facial es aquella tecnología que permite, a través de la IA, la identificación de personas y/o la comprobación de su identidad

[551] En inglés, EES.
[552] En inglés, ETIAS.
[553] Existiendo incluso un detector de identidades múltiples que verifica si los datos biográficos incluidos en la búsqueda se hallan en otros sistemas, con el fin de detectar casos de identidades múltiples vinculadas al mismo conjunto de datos biométricos.
[554]Véase Consejo de la Union Europea, 2019.

mediante el análisis y, en su caso, la comparación de sus datos biométricos, en concreto, de su rostro, a saber: formas, tipos, proporciones de sus contornos y rasgos de este.

Dichos sistemas, por lo general, con tales fines, cuentan con grandes bases de datos que albergan ingentes cantidades de imágenes faciales que servirán para, posteriormente, ser cotejadas por un algoritmo, previamente entrenado, con aquellas que puedan resultar de interés.

Así, en el ámbito de la investigación criminal, los sistemas de reconocimiento facial almacenan en sus bases de datos multipilicidad de imágenes de rostros de distintas personas para ser comparadas, en caso de necesidad, con las imágenes de las caras de aquellos sobre los que se pretende investigar, que habrán sido codificadas de forma automática por un algoritmo (previamente entrenado) tras su introducción en el sistema.

Las principales técnicas que se emplean por los mencionados sistemas, que habitualmente usan tecnología de *Machine Learning* (y *Deep Learning*), históricamente han sido fundamentalmente dos, a saber: técnicas de reconocimiento facial en 2 Dimensiones (2D) y técnicas de reconocimiento facial en 3 Dimensiones (3D), si bien esta segunda es claramente más eficaz y precisa que la primera, habida cuenta de que incluye tecnología infrarroja y permite distinguir si lo que se está capturando es realmente un rostro o, por ejemplo, una fotografía de un rostro, por lo que impide y/o dificulta el uso de caretas o fotografías para falsificar las caras de las personas.[555] No obstante, en la actualidad han irrumpido con fuerza también las técnicas de análisis facial, que desde luego son prometedoras.

Así, mientras la tecnología 2D es muy limitada, ya que requiere que la persona que se pretende analizar mire directamente a la cámara para ser precisa y efectiva (y, aun así, en muchas ocasiones su eficacia se ve comprometida), la tecnología 3D, que captura la imagen de un rostro en tiempo real (que posteriormente se procesa a través de una base de datos) es capaz de detectar los datos biométricos desde diversos ángulos de visión y en diversas circunstancias ambientales, incluso en la oscuridad. Por su parte, la novedosa técnica del

[555] Véase Kimaldi, s.f..

análisis facial lleva a cabo un examen detallado de diversas características del rostro humano y del entorno (entre otras, la textura de la superficie), lo que permite detectar incluso atributos modificados.[556]

Es importante poner de manifiesto, no obstante, que a pesar de las múltiples ventajas y avances que puede aportar una tecnología tan sofisticada como la descrita, tal y como se afirma desde Interpol[557], esta se enfrenta a especiales y complejos desafíos que dificultan su tarea identificativa.

Así, por un lado, los sistemas de reconocimiento facial, a diferencia de lo que sucede, por ejemplo, con los sistemas de reconocimiento de huellas dactilares y ADN (que son rasgos identificadores que permanecen invariables durante toda la vida de la persona), tienen que afrontar los diversos cambios, tanto naturales como artificiales, que pueden sufrir las características fisiológicas de un individuo a lo largo de su existencia, entre otros, el envejecimiento, las deformaciones sufridas por accidentes varios, la cirugía estética o reparadora, etc.

Por otro lado, tales herramientas, para llevar a cabo con éxito y precisión su labor, requieren de la existencia de imágenes de calidad determinada[558], lo cual no siempre resulta fácil de conseguir. Y es que, para obtener tal tipo de material no solo se necesitan equipos técnicos potentes capaces de captar imágenes nítidas y completas, sino que además hace falta que la/s persona/s analizada/s se halle/n en una posición que permita al sistema detectar sus datos biométricos con precisión, lo cual es altamente complicado (aunque, afortunadamente, los estándares de calidad exigidos van siendo cada vez menores).

Asimismo, la posible baja calidad de los algoritmos empleados por los sistemas de IA, según el nivel de entrenamiento que hayan tenido y según la información con que cuenten las bases de datos, puede conducir a la aparción de sesgos o discriminaciones y otras

[556] Véase Facial Recognition Market Size & Trends Report, 2021.
[557] Véase más en Interpol, s.f..
[558] Según Interpol, lo ideal sería disponer de una fotografía de pasaporte conforme a la norma OACI, ya que es una imagen frontal completa de la persona con iluminación homogénea en el rostro y un fondo neutro.

vulneraciones de derechos humanos absolutamente inaceptables, tal y como luego veremos con más profundidad.

Y, finalmente, tal y como se expondrá asimismo más adelante con mayor detalle, la introducción y el procesamiento de imágenes en las bases de datos de tales sistemas (cuya variedad, calidad y volumen son directamente proporcionales al éxito de su labor identificadora) entraña cuestiones jurídicas altamente delicadas, habida cuenta de la especial sensibilidad de los datos biométricos y, por ende, su elevada protección legal.

No en vano, y a pesar de los enormes avances alcanzados durante los últimos años, los expertos coinciden al afirmar que la tecnología de reconocimiento facial está todavía en ciernes y que queda un gran camino por recorrer.[559] Y, en relación con ello, la regulación de tal tipo de sistemas también se halla todavía en un momento bastante embrionario (especialmente en el ámbito español y de la UE), si bien, como veremos, ya se van sentando las bases de lo que será, sin duda, una de las temáticas estrella de los próximos años.

Y es que, ante una época de euforia inicial en la que parecía, con la irrupción con fuerza de las técnicas de reconocimiento facial, que estas iban a convertirse en un avance meteórico y muy eficaz, especialmente en materia de seguridad, en la actualidad, tras haberse detectado diversos fallos graves y haber quedado patente la falta de madurez de los sistemas (que, como se ha dicho, aun tienen un largo camino por delante para llegar a ser suficientemente fiables y precisos), países de todo el mundo están dando un paso atrás en relación a la autorización de su uso, cargándose de cautelas.

Así, por ejemplo, por un lado, en el ámbito de la UE, la Comisión Europea planteó en el documento publicado en 2019 como borrador del Libro Blanco de IA, titulado "*Structure for the White Paper on Artificial Intelligence-A European Approach*"[560], la prohibición del uso de técnicas de reconocimiento facial en lugares públicos por un periodo de entre tres y cinco años, con la finalidad de analizar los potenciales riesgos para los derechos

[559] Entre otros, un estudio realizado por la Universidad Carnegie Mellon de Pittsburgh (EEUU), presentó unas gafas diseñadas para evitar la identificación por parte de los sistemas de reconocimiento facial y demostró la vulnerabilidad de estos. En relación con ello, véase Hern, 2016.
[560] Véase Comision Europea, 2020.

fundamentales que estas entrañaban antes de dar luz verde a su uso masivo, si bien, finalmente, en el Libro Blanco de IA definitivo, publicado el 19 de febrero de 2020, se retiró tal recomendación y se dejó la decisión en manos de cada uno de los Estados Miembros. Y, por su parte, la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, tal y como se expondrá con más detalle más adelante, prevé su uso pero de forma limitada.

Y, por otro lado, en mayo de 2020, en EEUU, San Francisco se convirtió en la primera gran ciudad de tal país (a la que han seguido Oakland, Somerville y Boston, entre otras) en prohibir la utilización de sistemas de reconocimiento facial en la vía pública[561] bajo la premisa de que debe existir una regulación específica y garantista antes de que sean empleados de forma generalizada, especialmente por los poderes públicos (y, en concreto, por la policía para identificar delincuentes), para minimizar y/o evitar así los peligros que su uso puede entrañar para los derechos de los ciudadanos.

Y es que el descontrol y el limbo jurídico que en muchos países del mundo rodean a la tecnología analizada no hacen más que dejar la puerta abierta a constantes abusos y vulneraciones masivas de derechos fundamentales tanto por parte de los organismos públicos como de los entes privados que la emplean, lo cual resulta inaceptable.

En relación con ello, a modo de ejemplo, a principios de 2019 se hizo público que la agencia de Inmigración y Control de Aduanas de EEUU ("The Immigration and Customs Enforcement Agency of the United States" -ICE-) utilizó el reconocimiento facial para nutrir las bases de datos de los permisos de conducir estatales sin obtener previamente el consentimiento de los conductores.[562] Asimismo, el gobierno chino, tal y como veremos posteriormente con más profundidad, campa a sus anchas sin límite real alguno con el uso masivo de sistemas de reconocimiento facial en la vía pública, sin reparar en las posibles vulneraciones de derechos humanos que ello puede implicar para sus ciudadanos; y, en Rusia, las autoridades planean expandir el uso generalizado en espacios públicos de

[561] Véase Conger, Fausset & Kovaleski, 2019.
[562] Véase Harwell, 2019.

cámaras con *software* de reconocimiento facial a pesar de la falta de regulación, supervisión y garantía de protección de datos, tal y como ya ha denunciado la organanización Human Rights Watch.[563]

En la UE y en España, no obstante, la introducción de sistemas de reconocimiento facial con IA está empezando a proliferar con enormes cautelas, tal y como ya se ha advertido con anterioridad.

Así, por un lado, en el ámbito de la UE, existen diversos sistemas (o, en su caso, proyectos de sistemas) automáticos de identificación, que hacen uso de los datos biométricos faciales, entre los que resulta interesante mencionar: SIS II, VIS y EURODAC (que se analizará con más detalle posteriormente), por una parte; y SES[564], SEIAV[565] y ECRIS-TCN, por otra parte

En primer lugar, hay que hacer referencia al sistema SIS II ("*Schengen Information System second generation*"), gestionado por eu-LISA[566] y regulado por los siguientes instrumentos legales: Reglamento (UE) 2018/1860 del Parlamento Europeo y del Consejo, de 28 de noviembre de 2018, sobre la utilización del Sistema de Información de Schengen para el retorno de nacionales de terceros países en situación irregular; Reglamento (UE) 2018/1861 del Parlamento Europeo y del Consejo, de 28 de noviembre de 2018, relativo al establecimiento, funcionamiento y utilización del Sistema de Información de Schengen (SIS) en el ámbito de las inspecciones fronterizas, por el que se modifica el Convenio de aplicación del Acuerdo de Schengen y se modifica y deroga el Reglamento (CE) n.° 1987/2006; y Reglamento (UE) 2018/1862 del Parlamento Europeo y del Consejo, de 28 de noviembre de 2018, relativo al establecimiento, funcionamiento y utilización del Sistema de Información de Schengen (SIS) en el ámbito de la cooperación policial y de la cooperación judicial en materia penal, por el que se modifica y deroga la Decisión 2007/533/JAI del Consejo, y se derogan el Reglamento (CE) n.° 1986/2006 del Parlamento Europeo y del Consejo y la Decisión 2010/261/UE de la Comisión.

[563] Véase Human Rights Watch, 2020.
[564] En inglés, EES.
[565] En inglés, ETIAS.
[566] La Agencia de la Unión Europea para la gestión operativa de sistemas informáticos a gran escala en el espacio de libertad, seguridad y justicia.

Y es que el SIS II es un sistema informático de grandes dimensiones (cuyo predecesor, el SIS, fue creado en virtud de lo dispuesto en el Título IV del Convenio de aplicación del Acuerdo de Schengen de 14 de junio de 1985 entre los Gobiernos de los Estados de la Unión Económica Benelux, de la República Federal de Alemania y de la República Francesa, relativo a la supresión gradual de los controles en las fronteras comunes, más conocido como "Convenio de aplicación del Acuerdo de Schengen") que se ha convertido en "*el instrumento de intercambio de información más importante de Europa para garantizar la seguridad y una gestión eficaz de las fronteras*"[567], y que se constituyó como una medida de compensación a la libertad de circulación y a la eliminación de fronteras establecidas entre tales países, con el fin de contribuir a garantizar el mantenimiento de un alto nivel de seguridad, libertad y justicia en estos, confiriendo apoyo a "*la cooperación operativa entre las autoridades nacionales competentes, en particular guardias de fronteras, servicios policiales, autoridades aduaneras, autoridades de inmigración y autoridades responsables de la prevención, la detección, la investigación o el enjuiciamiento de infracciones penales o la ejecución de sanciones penales*."[568]

Y es que, el SIS II dispone de un sistema central (SIS II central), un sistema nacional en cada Estado Miembro ("N.SIS II") y una infraestructura de comunicación entre el sistema central y los sistemas nacionales[569], que permiten a las autoridades competentes de los mencionados Estados miembros consultar, introducir, actualizar o eliminar[570] información a través de sus sistemas locales.

Además, las autoridades competentes de cada país miembro podrán (previa valoración de la relevancia y las circunstancias del asunto, conforme a los artículos 21 y siguientes del Reglamento (UE) 2018/1862 del Parlamento Europeo y del Consejo de 28 de noviembre

[567] Considerando 8 del Reglamento (UE) 2018/1862 del Parlamento Europeo y del Consejo, de 28 de noviembre de 2018.
[568] Considerando 1 del Reglamento (UE) 2018/1862 del Parlamento Europeo y del Consejo, de 28 de noviembre de 2018.
[569] Véase Schengen information system second generation (SIS II), s.f..
[570] No obstante, las alertas sobre personas se eliminan de forma automática después de un período de tres años con posibilidad de ampliación., y las alertas sobre objetos se eliminan también automáticamente tras un período de cinco a diez años también con posibilidad de ampliación. según sean vehículos, barcos, aviones y contenedores u objetos que deban ser incautados o utilizados como pruebas en un procedimiento penal, respectivamente.

de 2018) generar "alertas" que contengan la descripción de una persona o un objeto con alguna de las finalidades siguientes: búsqueda y detención de individuos sobre los que pesa una orden de detención europea, localización de personas desaparecidas, evitación del traspaso de fronteras de personas con prohibición de entrada en el espacio Schengen, búsqueda de objetos robados, localización de menores con riesgo de sustracción, etc.

Las mencionadas "alertas", con tales fines, contendrán descripciones de personas (entre las que se incluyen sus datos estrictamente personales, tales como nombre y apellidos, género, edad y nacionalidad, y biométricos, tales como imágenes faciales, huellas dactilares y muestras de ADN) y de objetos, con indicación del motivo de la emisión de tal aviso.

No obstante, el Reglamento (UE) 2018/1862, del Parlamento Europeo y del Consejo, de 28 de noviembre de 2018, establece en su Capítulo XII unas normas específicas para datos biométricos, exigiendo un plus de calidad y control.

En concreto, el artículo 42, respecto de la introducción de fotografías e imágenes faciales, dispone que "*Solo se introducirán en el SIS las fotografías, las imágenes faciales* (…) *que cumplan las normas mínimas de calidad de los datos y las especificaciones técnicas. Antes de introducir ese tipo de datos, se comprobará su calidad, para determinar si se han cumplido las normas mínimas de calidad de los datos y las especificaciones técnicas*."

Y, por su parte, el artículo 43, en relación a la comprobación o consulta mediante fotografías e imágenes faciales establece en su apartado 1 que: "*Cuando una descripción del SIS disponga de fotografías, imágenes faciales* (…) *esas fotografías, imágenes faciales* (…) *se utilizarán para confirmar la identidad de una persona que haya sido localizada como consecuencia de una consulta alfanumérica realizada en el SIS*.", y añade en su apartado 4: "*Tan pronto como sea técnicamente posible, y siempre que se garantice un alto nivel de fiabilidad de la identificación, podrán utilizarse fotografías e imágenes faciales para identificar a una persona en el contexto de los pasos fronterizos oficiales.*
Antes de que se incorpore esta función en el SIS, la Comisión presentará un informe en el que examine si la tecnología necesaria está disponible, es fiable y está lista para su uso.
Se consultará al Parlamento Europeo sobre este informe.

Una vez que se haya empezado a utilizar esta función en los pasos fronterizos oficiales, la Comisión estará facultada para adoptar actos delegados con arreglo al artículo 75 a fin de completar el presente Reglamento en lo que se refiere a la determinación de otras circunstancias en las que se puedan utilizar fotografías e imágenes faciales para la identificación de personas."

Es importante remarcar que, además, cada Estado miembro que opera con el mencionado sistema dispone de una oficina SIRENE ("*Supplementary Information Request at the National Entry*")[571], que opera sin descanso y que se encarga de llevar a cabo la coordinación de toda aquella actividad relacionada con las alertas emitidas y los intercambios de información complementaria entre Estados miembros de forma estandarizada y segura.[572]

En segundo lugar, procede hacer mención al sistema VIS ("*Visa Information System*"), regulado por el Reglamento (CE) nº767/2008 del Parlamento Europeo y del Consejo, de 9 de julio de 2008, sobre el Sistema de Información de Visados (VIS) y el intercambio de datos sobre visados de corta duración entre los Estados miembros (Reglamento VIS) -cuya reforma (con mejora de los estándares de seguridad y protección de datos) está prevista para el año 2023[573]-, que permite a los Estados Schengen intercambiar datos de visados, facilitar el control y la emisión de tales documentos, controlar abusos (tipo "*visa shopping*"), proteger a los viajeros, ayudar con las solicitudes de asilo y reforzar la seguridad.

Y es que tal herramienta, gestionada también por eu-LISA, cuenta con un sistema central que, a través de una red de comunicaciones, conecta con los sistemas nacionales y todos los pasos fronterizos exteriores de los Estados Schengen (incluido Dinamarca) y procesa datos y decisiones relacionadas con las solicitudes de visados de corta duración para visitar o transitar por dicho espacio, pudiendo realizar una comparación biométrica, entre otras con imágenes faciales, con fines de identificación y verificación.

[571] En español, "*Solicitud de Información Complementaria a la Entrada Nacional*".
[572] Véase más en Comisión Europea, s.f..
[573] Véase más en Parlamento Europeo, s.f..

Con tal objetivo, se recogen 10 huellas dactilares y una fotografía digital de cada persona que solicita un visado y, tales datos, junto con los proporcionados en el formulario de solicitud de visa, se registran en una base de datos central segura, de forma que en las fronteras del espacio Schengen los datos biométricos del individuo que presenta el visado son comparados, mediante un escáner, con los que se hallan registrados en la base de datos.[574] Es importante poner de manifiesto, en relación con ello, que en casos específicos, las autoridades nacionales y Europol pueden solicitar el acceso a los datos introducidos en dicho sistema con el fin de prevenir, detectar e investigar delitos terroristas y otros ilícitos penales graves, tales como el tráfico de personas o de drogas, en los términos previstos en la Decisión 2008/633/JAI del Consejo, de 23 de junio de 2008, sobre el acceso para consultar el Sistema de Información de Visados (VIS) por las autoridades designadas de los Estados miembros y por Europol, con fines de prevención, detección e investigación de delitos de terrorismo y otros delitos graves.

En tercer lugar, procede hacer especial mención al denominado SES (Sistema de Entradas y Salidas de la Unión Europea)[575], desarrollado y gestionado por eu-LISA y regulado, por un lado, por el Reglamento (UE) 2017/2226 del Parlamento Europeo y del Consejo, de 30 de noviembre de 2017, por el que se establece un Sistema de Entradas y Salidas (SES) para registrar los datos de entrada y salida y de denegación de entrada relativos a nacionales de terceros países que crucen las fronteras exteriores de los Estados miembros, se determinan las condiciones de acceso al SES con fines policiales y, asimismo, se modifican el Convenio de aplicación del Acuerdo de Schengen y los Reglamentos (CE) nº767/2008 y (UE) nº1077/2011; y, por otro lado, por el Reglamento (UE) nº2017/2225 del Parlamento Europeo y del Consejo, de 30 de noviembre de 2017, por el que se modifica el Reglamento (UE) nº2016/399 en lo que respecta a la utilización del Sistema de Entradas y Salidas, ambos en vigor desde el 29 de diciembre de 2017.

Y es que tal sistema, previsiblemente en funcionamiento a partir de la primera mitad de 2022, tiene como finalidad sustituir al tradicional sistema de sellado manual de pasaportes, que resulta altamente ineficaz (puesto que conlleva mucho tiempo y no detecta muchas circunstancias que resultan relevantes, especialmente en el ámbito de la lucha contra el

[574] Véase más en Comisión Europea, s.f.
[575] En inglés, EES "*Entry/Exit System*".

terrorismo y otros delitos graves) y su principal objetivo es mejorar la eficiencia y la calidad de los controles realizados a aquellos viajeros que cruzan las fronteras exteriores del espacio Schengen para llevar a cabo estancias de corta duración. Para tales fines, el SES recogerá y almacenará en una base de datos central información relativa a la identidad de los viajeros de terceros países (incluidos sus datos biométricos, en concreto rasgos faciales y huellas dactilares) y de sus documentos de viaje, y tal información permitirá llevar a cabo la verificación automática de su identidad y de la autenticidad y la validez del documento oficial que presenten para cruzar las fronteras. Además, tal base de datos podrá ser consultada por Europol y por las Fuerzas y Cuerpos de Seguridad de los países miembros con fines de prevención, detección e investigación de delitos graves (incluido terrorismo), lo cual sin duda resultará muy útil y beneficioso en materia de seguridad.

En cuarto lugar, ha de hacerse referencia al programa Sistema Europeo de Información y Autorización de Viajes (SEIAV)[576], regulado, por una parte, por el Reglamento (UE) nº2018/1240 del Parlamento Europeo y del Consejo, de 12 de septiembre de 2018, por el que se establece un Sistema Europeo de Información y Autorización de Viajes (SEIAV) y por el que se modifican los Reglamentos (UE) nº1077/2011, (UE) nº515/2014, (UE) nº2016/399, (UE) nº2016/1624 y (UE) nº2017/2226; y, por otra parte, por el Reglamento (UE) nº2018/1241 del Parlamento Europeo y del Consejo, de 12 de septiembre de 2018, por el que se modifica el Reglamento (UE) nº2016/794 con objeto de establecer el Sistema Europeo de Información y Autorización de Viajes (SEIAV).

Y es que tal sistema, asimismo desarrollado y gestionado por eu-LISA, que estará operativo a finales de 2022, tiene como objeto la expedición de una autorización de viaje a todos aquellos ciudadanos de países que no necesitan visado para acceder a cualquiera de los veintiséis Estados del espacio Schengen, con el fin de proteger y fortalecer sus fronteras (al igual que ocurre con la autorización de viaje ESTA para ingresar en EEUU).

Si bien tal autorización de viaje ETIAS se tramitará *on line*, los solicitantes deberán aportar su nombre completo, fecha de nacimiento, país de residencia, detalles de un pasaporte válido, dirección de correo electrónico y tarjeta de débito o crédito para abonar la tarifa de

[576] En inglés, "*European Travel Information and Authorization System*" (ETIAS).

la solicitud y, en cuestión de minutos, tal información será cotejada con una serie de bases de datos de seguridad, entre ellas, las de SIS, VIS, EUROPOL DATA e Interpol, que como veremos contienen datos biométricos (huellas dactilares y rasgos faciales), y posteriormente el sistema emitirá una resolución de autorización favorable o desfavorable en función de los hallazgos obtenidos.[577]

En quinto lugar, procede hacer alusión al Servicio Europeo de Información de Antecedentes Penales: Información sobre condenas de nacionales de terceros países (ECRIS-TCN), regulado por el Reglamento (UE) nº2019/816 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, en vigor desde el 11 de junio de tal año, por el que se modifica el Reglamento (UE) 2018/1726 y se establece un sistema centralizado para la identificación de los Estados miembros que poseen información sobre condenas de nacionales de terceros países y apátridas (ECRIS-TCN) a fin de complementar el Sistema Europeo de Información de Antecedentes Penales.

En virtud de tal sistema, las autoridades nacionales tienen la obligación de crear un registro de datos de cada nacional de terceros países al que condenen, debiendo incluir sus datos identificativos, entre otros, los datos biométricos, a saber, la información dactiloscópica que se haya recogido conforme al Derecho nacional del Estado miembro en cuestión durante los procesos penales, y las imágenes faciales siempre y cuando tal legislación lo permita. Tras ello, dichos datos deben ser introducidos en el sistema central de ECRIS-TCN, lo que permite a las autoridades nacionales consultar y recabar información sobre los antecedentes penales de un individuo, teniendo asimismo acceso directo a tal gran base de datos Eurojust, Europol y la Fiscalía Europea para identificar a aquellos Estados miembros de la UE que poseen información sobre condenas penales de nacionales de países de terceros Estados.

Y, finalmente, procede hacer especial mención a los planes de cambio y a la discusión que existen alrededor de la posibilidad de incluir datos biométricos consistentes en imágenes faciales en la base de datos creada en virtud de la Decisión 2008/615/JAI del Consejo, de 23 de junio de 2008, sobre la profundización de la cooperación transfronteriza, en particular

[577] Véase European Travel Information and Authorization System (ETIAS), s.f..

en materia de lucha contra el terrorismo y la delincuencia transfronteriza (informalmente conocida como "Decisión Prüm"), accesible por los cuerpos policiales de los países firmantes de tal instrumento legal, para fines de investigación criminal (principalmente por delitos de terrorismo y delincuencia transfronteriza).

Y es que el 5 de julio del 2018, el Consejo de la Unión Europea, en el marco del "*Proyecto de Conclusiones del Consejo sobre la aplicación de las «DECISIONES PRÜM» diez años después de su adopción*", publicó un documento de recomendaciones a los Estados Miembros en que les sugirió que invitaran al grupo de expertos DAPIX (Grupo de Trabajo sobre el intercambio de información y protección de datos) a evaluar la evolución de los trabajos de Prüm para introducir más avances "*con vistas a utilizar posibles tecnologías biométricas novedosas, como los sistemas de reconocimiento facial*"[578], lo cual suscitó inquietud y dudas jurídicas en algunos miembros del Parlamento Europeo, en concreto de la Comisión de Libertades Civiles, que alertaron del posible abuso del uso de datos y de las posibilidades de falsos positivos.[579]

Por otro lado, en concreto, en el ámbito de España (al igual que en otros países de la UE), actualmente algunos puntos fronterizos -especialmente en aeropuertos, tales como el de Alicante, Barcelona, Gerona, Ibiza, Mallorca, Málaga y Madrid, entre otros; y puertos, tales como el de Algeciras- cuentan con máquinas denominadas "*Automated Border Control*" (ABC) que, cada vez que un individuo introduce un pasaporte para su validación, remiten su información a todas las bases de datos europeas, lo que permite detectar de forma prácticamente inmediata si aquél tiene alguna causa pendiente o prohibición en el ámbito del espacio Schengen (por ejemplo, una orden de busca y captura, una prohibición de entrada/salida del país, etc), lo que hace saltar una alarma inmediata que conecta directamente con Policía Nacional, que se encarga de gestionar el asunto *in situ*.

Sin embargo, la información remitida y empleada para los mencionados fines todavía no es biométrica, sino que contiene simplemente los datos del pasaporte analizado. Y es que en tales sistemas, los datos biométricos (rostro y huellas dactilares) solo se utilizan para verificar que la persona que introduce el pasaporte en la máquina es la misma que su titular,

[578] Véase texto completo en Consejo Europeo, 2018.
[579] Véase Stolton, 2020.

a los efectos de detectar posibles usurpaciones de identidad o usos de documentación falsa. No obstante, las máquinas ABC registran los datos personales y de viaje de los usuarios, lo que resulta de gran valor para potenciales investigaciones, que pueden contar con información valiosísima sobre los movimientos de presuntos delincuentes y/o víctimas. Lo que no hacen todavía, sin embargo, es conservar sus datos biométricos e introducirlos en una base de datos, a pesar de que sin duda podría resultar de gran utilidad.

No obstante, lo anterior se limita a los ciudadanos del espacio Schengen, si bien con la puesta en marcha del anteriormente mencionado sistema SES[580] (previsiblemente durante la primera mitad de 2022), el control de pasaportes mediante los datos biométricos de los viajeros se llevará a cabo con ciudadanos de terceros países -fuera del espacio Schengen-, que deberán dejar constancia en las denominadas cabinas biométricas de las huellas dactilares de los cuatro dedos de la mano derecha (excepto el pulgar) y de los rasgos de la cara, que se introducirán en una base de datos que permitirá (o denegará) la entrada y salida automática a los países del espacio Schengen durante tres años (ya que posteriormente se tendrán que actualizar).

En cualquier caso, si bien los mencionados sistemas pueden proporcionar de forma automática información muy valiosa para la prevención y la investigación de delitos en cuestión de segundos, tal y como me aseguró Fernando Agoiz Rodríguez, Comisario Jefe del Puesto Fronterizo del aeropuerto de Barcelona -El Prat-, del Cuerpo Nacional de Policía, el papel del profesional humano seguirá siendo clave para tales fines. Y es que, al menos por el momento, la IA no es capaz todavía de detectar comportamientos o situaciones que, a un policía experimentado y comprometido, no se le pasan por alto, por lo que este siempre deberá estar detrás de tal tecnología, que sin duda le servirá como complemento para desarrollar su labor de un modo más eficiente y exitoso.

Así, cierto es que puede entrenarse a una "máquina" para que, en caso de que detecte, por ejemplo, el incremento de viajeros de una nacionalidad distinta a la del país de origen, haga saltar una alarma, si bien en la actualidad resulta impensable esperar que el sistema

[580] Véase más en Comisión Europea, s.f..

averigüe si ello tiene trascendencia criminal o no, lo cual solo podrá ser descifrado por profesionales "con olfato", veteranía y mucha profesionalidad.

Por su parte, la Secretaría de Estado de Serguridad (dependiente del Ministerio del Interior) ya ha adquirido un *software* de reconocimiento facial a la compañía francesa Thales para distribuir entre los cuerpos policiales de Policía Nacional, Guardia Civil, Mossos d'Esquadra y Policía Foral de Navarra. La idea que subyace bajo tal iniciativa es que las distintas Fuerzas de Seguridad del Estado vayan introduciendo en la base de datos del sistema las imágenes de los rostros de aquellas personas que resulten detenidas y reseñadas, con el fin de establecer un gran espacio de cooperación, que permitirá compartir información sin duda valiosísima para la investigación criminal, y crear nuevos mecanismos de identificación, si bien todavía el proyecto está muy en ciernes.

La Dirección General de la Guardia Civil, por su parte, tal y como se desprende de la Resolución de 16 de julio de 2020, de la Subsecretaría, por la que se publica el Convenio entre el Centro para el Desarrollo Tecnológico Industrial, E.P.E., y el Ministerio del Interior, relativo a la contratación precomercial de servicios de I+D en materia de seguridad en el medio rural, ha suscrito un Convenio con el Centro para el Desarrollo Tecnológico Industrial, E.P.E., relativo a la contratación precomercial de servicios de I+D, en materia de seguridad en el medio rural, en cuya virtud se prevé la utilización de sistemas de reconocimiento facial para el control de eventos multitudinarios.

En concreto, en dicho Convenio se hace especial mención a un popular evento masivo, el festival Viña Rock, que suele celebrarse en Villarrobledo (Albacete), y se refiere al uso, entre otros, de un sistema de reconocimiento facial que debería ser instalado en el punto de control de acceso al mismo con el objetivo de proporcionar a los agentes alertas para detener a aquellos individuos que tengan asuntos pendientes con la justicia, lo cual resulta altamente útil, especialmente en los términos del mencionado Convenio, que establece el compromiso de cumplir lo dispuesto en el RGPD.

a.2) Posibles utilidades en la instrucción de las causas

Como puede ser intuido, las tecnologías de reconocimiento facial tienen infinidad de utilidades tanto en el sector privado como en el sector público, a saber, entre otras, la identificación del rostro de los trabajadores de una compañía para comprobar sus entradas y salidas del trabajo; la detección por parte de aplicaciones de citas, como por ejemplo Tinder, de caras consideradas atractivas para hacer coincidir sus perfiles con los de personas que, según su algoritmo, son igualmente sugerentes; o la identificación del rostro de los ciudadanos nacionales de un país en máquinas dispuestas en los servicios de fronteras en ciertos puertos y aeropuertos para agilizar los trámites de llegada y partida, tal y como se ha visto.

El enorme potencial que tiene la aplicación de los mencionados sistemas, pues, sin lugar a duda puede resultar de gran impacto en muchos ámbitos, pero procede centrar el foco en el de la investigación criminal, que es el que da razón a la presente tesis doctoral.

Así, lo que pretendo poner de manifiesto a continuación no es más que una relación de los distintos usos y utilidades que las herramientas de IA de reconocimiento facial, en mi opinión, pueden tener en el proceso de instrucción español, en aras de detectar y mostrar los enormes beneficios que pueden aportar (principalmente para la identificación del/los autor/es de los hechos delictivos), que posteriormente, no obstante, serán puestos en contrapeso con los posibles riesgos, especialmente jurídicos, que pueden entrañar.

A pesar de lo expuesto, no obstante, procede advertir que la utilidad del uso de herramientas de IA de reconocimiento facial mayoritariamente dependerá de la existencia de imágenes dubitadas (de calidad) de las personas de interés para que puedan ser comparadas con las imágenes indubitadas que consten en las correspondientes bases de datos policiales y/o judiciales, por lo que su potencial es limitado, al menos por el momento, siendo que España no cuenta con sistemas de videovigilancia masiva. No obstante, cierto es que cada vez existen más posibilidades de lograr imágenes a través de las cámaras de los teléfonos móviles de los ciudadanos, de las cámaras de seguridad colocadas en la vía pública o en establecimientos privados, etc, por lo que resulta muy interesante explorar las posibilidades que tal tecnología ofrece, ya que en un ámbito como el de la investigación criminal toda ayuda es poca y, estoy segura, además, de que las posibles utilidades irán creciendo

exponencialmente en los próximos años hasta niveles insospechados (aunque deberán implementarse límites jurídicos claros).

Así, en China, por ejemplo, donde hay desplegadas más de ciento setenta millones de cámaras de videovigilancia por todo el país, la localización de un sospechoso cuya imagen ha sido previamente introducida en su base de datos tiene una media de duración de siete minutos, tal y como se demostró por el reportero de la BBC John Sudworth, que cedió una imagen de su rostro a la policía china y comprobó que esta tardaba tal lapso de tiempo en darle el alto mientras caminaba por la vía pública, lo cual es absolutamente abrumador.[581] Y en Rusia van por el mismo camino, ya que en enero de 2020 se hizo público que la ciudad de Moscú iba a empezar a usar cámaras de reconocimiento facial proporcionadas por la compañía NtechLab que notificarían a la policía de forma inmediata y en directo los eventuales hallazgos de personas etiquetadas como sospechosas en su base de datos.[582]

Ante tales utilidades, pues, cada vez más compañías, entre otras Ayonix Corporation, por ejemplo, se están especializando en crear sistemas de reconocimiento facial en tiempo real con el uso de cámaras IP ("*Internet Protocol camera*") y en comercializarlos entre los distintos cuerpos de policía del mundo entero para ayudar a identificar sospechosos en lugares públicos en tiempo récord.

Empezamos.

a.2.1) Sustitución o complemento de la diligencia de investigación consistente en la rueda de reconocimiento en aquellos casos en que haya testigos

El Capítulo 3º del Título 5º de la LECrim lleva por título "De la identidad del delincuente y de sus circunstancias personales" y contiene los artículos 368 y siguientes, que regulan la diligencia de investigación consistente en la rueda de reconocimiento.

En relación con ello, el artículo 368 LECrim dispone:

[581] Véase BBC News, 2017.
[582] Véase Vincent, 2020.

"*Cuantos dirijan cargo a determinada persona deberán reconocerla judicialmente, si el Juez instructor, los acusadores o el mismo inculpado conceptúan fundadamente precisa la diligencia para la identificación de este último, con relación a los designantes, a fin de que no ofrezca duda quién es la persona a que aquéllos se refieren*."

Y, por su parte, el artículo 368 LECrim establece:

"*La diligencia de reconocimiento se practicará poniendo a la vista del que hubiere de verificarlo la persona que haya de ser reconocida, haciéndola comparecer en unión con otras de circunstancias exteriores semejantes. A presencia de todas ellas, o desde un punto en que no pudiere ser visto, según al Juez pareciere más conveniente, el que deba practicar el reconocimiento manifestará si se encuentra en la rueda o grupo la persona a quien hubiese hecho referencia en sus declaraciones, designándola, en caso afirmativo, clara y determinadamente*.
En la diligencia que se extienda se harán constar todas las circunstancias del acto, así como los nombres de todos los que hubiesen formado la rueda o grupo."

En relación con ello, la jurisprudencia ha señalado de forma reiterada que la rueda de reconocimiento es "*una diligencia sumarial que tiene por fin la identificación del inculpado en cuanto sujeto pasivo del proceso, de manera que para que produzca efectos probatorios es imprescindible, como regla general, que la diligencia practicada ante el juez de instrucción con las formalidades y garantías legalmente previstas sea posteriormente ratificada en el juicio oral por el testigo que hizo el reconocimiento, para que esta declaración pueda ser sometida a contradicción con sujeción a los principios de oralidad e inmediación, como exigen las garantías constitucionales inherentes al justo proceso (entre otras, SSTC 205/1998, 164/1998, 148/1996, 32/1995 y 283/1994; SSTS 930/2013, 601/2013, 428/2013 y 503/2008)*."[583]

Así, de lo expuesto se deduce que "*la necesidad de su práctica surge fundamentalmente en aquellos supuestos delictivos en que, por no existir relaciones previas entre el autor del delito y la víctima, ésta no pueda proporcionar a los investigadores los datos a que se*

[583] García, 2014, pág. 8.

refiere el art. 277.3 LECrim , o cualesquiera otros (alias, mote, apodo, sobrenombre, parentesco, paradero profesional, etc...) que sirvan al mismo fin (arts. 142.1 y 388 LECrim). (…) *El reconocimiento en rueda es una diligencia esencial pero no inexcusable, supone un medio de identificación, no exclusivo ni excluyente y así el art. 369 LECrim, parte de que sea precisa por las circunstancias concurrentes ofrezca duda de identificación y la omisión del reconocimiento en rueda no significa por sí misma, la vulneración de ningún precepto constitucional*", tal y como se dispone en la STS 786/2017, de 30 de noviembre (Sala 2ª) FJ 1º.

Y es que en la práctica, la dinámica de dicha diligencia de investigación resulta clara: el juez de instrucción la acuerda cuando la considere necesaria e indispensable para la identificación del/los autor/es de los hechos y, tras ello, se cita al/los investigado/s al juzgado para someterlo/s, junto con otros figurantes de características físicas semejantes a las suyas (normalmente un total de cinco candidatos, aunque el mínimo son tres), a identificación por parte de la/s persona/s que tuvieron contacto visual con el/los autor/es de los hechos (por lo general, la propia víctima o un testigo), que deberá/n manifestar al juez (previa advertencia por parte de este de que dicho autor puede estar o no entre los figurantes), en presencia de los Letrados de la defensa y, en su caso, la acusación, el Ministerio Fiscal y el/la Letrado/a de la Administración de Justicia, que da fe, si reconoce/n a alguno de ellos y, si es así, en qué porcentaje de seguridad, de lo cual quedará constancia en un acta firmada por todos los presentes. Posteriormente, en caso de que el procedimiento siga adelante, tal acta de la diligencia practicada se remite al órgano enjuiciador para ser sometida a contradicción en el acto del plenario y, posteriormente, ser valorada como un elemento probatorio más en la sentencia que se dicte.

De acuerdo con lo expuesto, parece que nos hallamos ante una diligencia instructora fácil y efectiva, pero la realidad dista mucho (por no decir muchísimo) de ello.

Así, por un lado, es una problemática común a todos los juzgados de instrucción la enorme dificultad que existe a la hora de formar las ruedas de reconocimiento, puesto que encontrar a personas con características y rasgos físicos similares a los de los presuntos autores de los hechos, es tarea complicada.

Y es que son muchos los factores físicos de los investigados que deben tenerse en cuenta para conseguir formar una rueda de reconocimiento de calidad que luego tenga valor en el acto del plenario, a saber: la raza, los rasgos faciales, el color y la forma del cabello, el volumen corporal, la estatura, los accesorios (gafas, aparatos dentales, etc), marcas corporales, tatuajes, etc. Así, si bien la policía judicial suele realizar verdaderos esfuerzos (a pesar de sus limitados medios personales y materiales) para intentar hallar candidatos que puedan resultar idóneos para hacer de figurantes en las ruedas de reconocimiento (buscando principalmente voluntarios en la vía pública y en los centros penitenciarios), lo cierto es que es una tarea que resulta muy difícil de ejecutar.

Y tanto es así que la jurisprudencia del Tribunal Supremo (entre otras, STS 18/2017, de 20 de enero) ha interpretado que la exigencia legal de personas de características similares a las del investigado es un mero *desideratum* condicionado por la posibilidad de contar con individuos de circunstancias externas semejantes, advirtiendo de que tal exigencia no puede ser concebida de forma tan rigurosa que haga imposible la práctica de la diligencia.[584]

Como es lógico, no obstante, los Letrados de la defensa suelen ser muy estrictos y exigentes con la calidad de las ruedas de reconocimiento, habida cuenta de que en caso de que los figurantes no tengan un parecido sustancial con sus clientes, resultará más fácil su identificación (aunque solo sea por eliminación), por lo que tienden a impugnar tal diligencia que, mal hecha, sin duda podría vulnerar el derecho a la defensa de las personas investigadas.

Ademas de lo expuesto, muchas veces hay que lidiar con los cambios de última hora a los que se someten las personas investigadas, que aparecen el día señalado para la práctica de la rueda de reconocimiento con características físicas distintas a las que presuntamente tenían en el momento de los hechos, en ocasiones de forma involuntaria, simplemente a causa de los largos periodos de tiempo que transcurren desde tal instante hasta que se practica la diligencia, y a veces de forma deliberadamente voluntaria, *motu propio* o por

[584] Así, solo se considera que no concurren circunstancias semejantes cuando se presenten personas de diferente raza o sexo, pero se entiende bastante que los figurantes vistan en forma semejante y tengan estaturas y condiciones físicas no extremadamente diferentes.

consejo de su Letrado o de algún conocido. Así, por ejemplo, una mujer investigada por delito de estafa que en el momento de comisión de los hechos tenía una melena larga y rubia, aparatos dentales y ojos azules, puede aparecer el día de la rueda de reconocimiento con el pelo corto y negro, sin rastro del aparato dental y lentillas marrones, lo cual hace muy difícil, si no imposible, la tarea identificadora y, asimismo, puede provocar la suspensión de la diligencia, con todo lo que ello conlleva.

Por otro lado, bien es sabido que la justicia en este país, por desgracia, es lenta (en ocasiones, extremadamente lenta). Y ello no puede resultar más nocivo para el éxito de las diligencias de reconocimiento en rueda, habida cuenta de que el paso del tiempo es su principal enemigo, puesto que difumina recuerdos, facilita el cambio físico de los autores de los hechos, como se ha dicho y, en ocasiones, implica la sustracción de estos a la acción de la justicia, lo cual conlleva en muchos casos el archivo de las causas, la prescripción o la absolución de los acusados (en ocasiones, por delitos muy graves que quedan impunes).

Además, la identificación de los autores de los hechos por parte de las víctimas o los testigos, no siempre resulta tarea fácil. Y es que en algunas ocasiones nos encontramos con personas vulnerables (menores de edad, ciudadanos con sus capacidades intelectivas mermadas o de la tercera edad) a las que les cuesta muchísimo recordar a los presuntos delincuentes; en otras, nos topamos con personas que sufrieron graves situaciones de nervios y estrés en el momento de la comisión del delito y, simplemente, han borrado de su mente, a modo de autoprotección, los rasgos físicos de los autores; en otras, nos hallamos con gente a la que la práctica de la diligencia de reconocimiento en rueda le supone una angustia y un miedo tal que les impide llevar a cabo la tarea identificadora con claridad y eficacia, siendo incapaces en muchos casos de volver a mirar a la cara a los presuntos delincuentes; y, en otras tantas, las víctimas o testigos simplemente ven imposible recordar el rostro del delincuente habida cuenta de las circunstancias ambientales en que se produjeron los hechos, tales como, entre otras, las condiciones de luz, el lugar, la duración del suceso, el tiempo de exposición a la cara del autor, o la distancia entre este y la víctima o testigo o el número de agresores, tal y como indica la jurisprudencia (entre otras, SSTS 901/2014, de 30 de diciembre y 473/2016, de 1 de junio).

De acuerdo con ello, se han publicado múltiples estudios en el ámbito de la psicología del testimonio que advierten de la existencia (no poco frecuente) de identificaciones erróneas, incluso en aquellos casos en que las víctimas o testigos aseguran reconocer sin ningún género de dudas al/los autor/es de los hechos. Y es que los antedichos estudios muestran, por ejemplo, que los humanos tenemos más capacidad para reconocer a sujetos de nuestra propia raza, lo que aumenta el riesgo de identificación inexacta cuando se imputa la comisión criminal a ciudadanos pertenecientes a razas distintas; asimismo, concluyen que el grado de seguridad del testigo a la hora de realizar la identificación no es un indicador fiable de la precisión de la misma; y también aseveran que cuanto menor sea el número de figurantes de una rueda, mayor puede resultar la sugestión del sujeto identificador.[585]

A ello hay que añadir, además, la desagradable experiencia que supone para cualquier víctima o testigo acudir al juzgado a reconocer al autor de unos hechos que, con toda seguridad, no le habrá dejado indiferente. Y es que, si bien el grado de afectación depende de cada persona y de cada caso, lo cierto es que la práctica de la diligencia de reconocimiento en rueda en ocasiones genera verdaderos traumas a aquellos que deben proceder a la identificación, ya que por mucho que se les explique que no serán vistos por los figurantes (lo cual es cierto, aunque a veces hay que hacer verdaderos malabares para conseguirlo, debido a las pésimas y precarias instalaciones de los juzgados), ellas saben que en el expediente, donde consta su completa filiación, va quedar reflejado que procedieron a "señalar" y a reconocer al presunto autor de los hechos, lo cual siempre genera temor a posibles futuras represalias (más aun en aquellos casos en que en el propio acto del juicio se les pregunta además de forma explícita si reconocen a la persona que está sentada en el banquillo de los acusados y para contestar deben mirarla a la cara).

En adición, en ciertas ocasiones las víctimas o testigos tienen ciertas influencias externas que distorsionan la pureza de su labor identificadora. Así, por ejemplo, hay veces en que, en el periodo de tiempo que media entre el hecho delictivo y la práctica de la rueda de reconocimiento, tales víctimas o testigos reciben información externa de terceros sobre la identidad del delincuente, que les es mostrado por la policía en una fotografía o señalado por la calle por un amigo; y, hay veces, también, en que, a pesar de los esfuerzos realizados

[585] Véase De Paul, 2009, págs. 7-18.

por parte del personal de los juzgados para impedirlo, las víctimas o testigos se cruzan con el/los investigado/s de los hechos en sede judicial antes de la práctica de la rueda de reconocimiento, lo cual les predispone a señalarlo/s de forma directa, a pesar de que quizás no sean las personas que realmente cometieron el/los delito/s perseguido/s.

Tales circunstancias, desde luego, generan nefastas consecuencias no solo para la persona identificada, que puede resultar condenada siendo inocente, sino también para la sociedad, puesto que el verdadero autor de los hechos, que puede representar un gran peligro para la comunidad, podría quedar impune y, por ende, permanecer en libertad.

A todo ello, además, hay que añadir la especial dificultad que, para la tarea identificadora, surge en aquellos casos en que a la persona investigada no le resulte posible por algún motivo justificado acudir a la sede judicial para ser sometida a la práctica de la rueda de reconocimiento y esta deba realizarse por videoconferencia, puesto que, desde luego, la calidad de las imágenes mostradas a la víctima o testigo a través de una pantalla de televisión hacen muchísimo más ardua su labor de identificación que en aquellos casos en que el reconocimiento se hace a pocos centímetros de distancia y en directo, con la única separación de un cristal.

Finalmente, otras de las grandes complicaciones que las ruedas de reconocimiento presentan en la práctica son, por un lado, las más que frecuentes, por desgracia, inomparenecias de la/s persona/s investigada/s el día señalado para su realización, con el enorme desperdicio de recursos que ello conlleva, habida cuenta de la dificultad que, como se ha expuesto, entraña formar una rueda tanto para la policía judicial como para el personal del Juzgado (así como para los figurantes y los Letrados, que ven malgastado su tiempo); y, por otro lado, los también más que frecuentes casos en que el/los presunto/s autor/es de los hechos, a pesar de estar identificado/s, se halla/n en paradero desconocido y bajo orden de busca y captura. En estos últimos casos, generalmente, transcurre un tiempo absolutamente desproporcionado entre el momento de la comisión de los hechos y el de la práctica de la rueda de reconocimiento, lo que conlleva un mayor obstáculo para las víctimas o testigos que deben proceder a la identificación, tal y como ya se ha advertido con anterioridad, mucho tiempo después. Y es que, la persona investigada primero debe ser hallada (lo cual, en ocasiones, tarda años) y, luego, a pesar de que sea detenida y puesta

a disposición judicial (puede ser en cualquier Juzgado de Instrucción en servicio de guardia del territorio español, en caso de que se halle bajo orden de busca y captura nacional), deberá ser citada para comparecer ante aquel órgano judicial que tramita la causa (puesto que el Juzgado de guardia que recibe al detenido no suele tener posibilidad de formar una rueda de reconocimiento en el poco tiempo que dura su servicio y, además, no es deseable que se celebre por videoconferencia), lo cual no siempre tiene resultado de éxito. Asimismo, y aun en el caso de que la persona investigada decida comparecer ente tal órgano judicial, siendo que las agendas están llenas y siendo que, como se ha dicho, formar una rueda de reconocimiento no es tarea fácil, pasa mucho tiempo hasta que esta puede practicarse (tanto que, en múltiples ocasiones, las víctimas/testigos ya son incapaces de reconocer a los autores de los hechos).

En virtud de todo lo expuesto, quedan patentes las múltiples debilidades y carencias que alberga la práctica de los reconocimientos en rueda que se llevan a cabo actualmente en nuestros juzgados, que no hacen más que mermar su valor y utilidad como diligencia de investigación. Y es que, o bien su ejecución resulta imposible, o bien deviene poco exitosa, o bien, cuando se realiza con éxito, el valor que se le otorga por los órganos enjuiciadores como prueba de cargo no es el deseado por su cuestionada calidad, por lo que muchas veces surge la pregunta de si realmente compensa su práctica, puesto que, además, en numerosas ocasiones, a pesar de las enormes dificultades y el despliegue de recursos que estas conllevan, el mero hecho de que la víctima o testigo no logre reconocer a la persona investigada o lo haga con escasa seguridad (aunque sea por alguno de los motivos externos anteriormente expuestos), si no hay más pruebas de cargo contundentes, decanta la balanza hacia la absolución de personas culpables.

Respecto del valor probatorio que se otorga por los órganos enjuiciadores a la diligencia de reconocimiento en rueda practicada en fase de instrucción con todas las garantías, en la STS 289/2020, de 5 de junio (Sala 2ª) FJ 1º, se dispone:

"*c) El reconocimiento en rueda practicado ante el juez de instrucción para ser entendido como prueba válida y suficiente para desvirtuar la presunción de inocencia debe ser reproducido en el juicio oral mediante ratificación, a fin de poder ser sometida quien lo haya realizado a las garantías de contradicción, oralidad e inmediación. Es esencial que,*

siendo posible, la víctima o testigo acudan al plenario para ratificar dicha diligencia ya que, como prueba testifical, es, por su naturaleza, perfectamente reproducible en el acto del juicio oral y debe ser, por tanto, sometida a contraste y contradicción por las partes de forma oral y sin mengua de los derechos de defensa del imputado. Todo ello de conformidad con lo dispuesto en el art. 6.3 d) del Convenio Europeo de Derechos Humanos, que manifiesta que todo acusado tiene, entre sus mínimos derechos, el de "interrogar o hacer interrogar a los testigos que declaren contra él", así como con el art. 14.3 e) del Pacto Internacional de Derechos Civiles y Políticos, del mismo tenor.

d) El reconocimiento directo y sin ningún género de dudas ante el tribunal es prueba suficiente para enervar la presunción de inocencia a pesar de las irregularidades de los reconocimientos fotográficos, o incluso de reconocimientos en rueda anteriores. Esta Sala ha declarado en la que "cuando el testigo señala inequívocamente a una persona durante el plenario, su fuerza probatoria radica en la credibilidad o fiabilidad del testimonio de quien realiza la identificación" (STS nº 177/2003, de 5 de febrero). Un reconocimiento dudoso en fase sumarial puede ser subsanado mediante uno inequívoco en el Plenario o viceversa, cuando en la fase de instrucción se ha producido una rueda de reconocimiento con todas las formalidades legales y quien realiza el reconocimiento no ha admitido dudas sobre la identidad del reconocido y en el Plenario las suscita; el Tribunal, entonces previa introducción de dicha diligencia en el juicio oral, puede acoger la que le ofrezca mayor verosimilitud (STS 1278/2011, de 29 de noviembre)."

Y es que ello demuestra, pues, que la diligencia de investigación de reconocimiento en rueda, por sí misma, carece de valor probatorio válido y suficiente para desvirtuar la presunción de inocencia de la persona acusada, siendo que la jurisprudencia exige que la identificación realizada por la víctima o testigo en fase de instrucción sea reproducida en el acto del plenario mediante ratificación, para poder ser así sometida a contraste y contradicción por las partes de forma oral, en aras de garantizar el derecho a la defensa de la/s persona/s investigada/s. No obstante, de la jurisprudencia traída a colación se deduce que, por un lado, en caso de que la víctima o testigo hiciera un reconocimiento dudoso en fase de instrucción, este podría ser subsanado si en el acto del plenario procediera al reconocimiento inequívoco del presunto autor de los hechos; y, al revés, por otro lado, se

podría otorgar fuerza al resultado indubitado de la diligencia de reconocimiento en rueda si posteriormente a la víctima o testigo le surgieran dudas en el acto del juicio.

En cualquier caso, los órganos enjuiciadores, conscientes de las limitaciones y vulnerabilidades que presentan las identificaciones visuales humanas, suelen ser muy cautelosos a la hora de valorar los resultados de las diligencias de reconocimiento en rueda practicadas en fase de instrucción y sus posteriores ratificaciones realizadas en el acto del plenario, especialmente cuando se trata de fundamentar una sentencia condenatoria. Y es que, en virtud de ello, tales órganos tienden a valorar siempre todas las particularidades concurrentes en el caso concreto para poder así determinar la fiabilidad de los reconocimientos efectuados, teniendo en cuenta tanto las circunstancias existentes en el momento de la comisión del delito, como las concurrentes en el momento de la identificación realizada en fase de instrucción, y las presentes en la declaración testifical o ratificación del reconocimiento en rueda en el acto del juicio oral.

En relación con todo lo anterior, es interesante, además, hacer especial y breve referencia a los denominados reconocimientos fotográficos que se llevan a cabo, en muchas ocasiones, en sede policial, por su estrecha vinculación con las diligencias de reconocimiento en rueda.

Y es que los reconocimientos fotográficos policiales, por lo general, se realizan, o bien en aquellos casos en que la víctima o testigo, sin relación previa con el/los autor/es de los hechos, hace una descripción del/los mismo/s que induce a la policía a pensar en una/s persona/s concreta/s, cuya/s imagen/es le/s deberá/n ser mostrada/s para su identificación, siempre junto con las imágenes de otras personas con carácterstícas físicas similares, en aras de no viciar ni tal diligencia ni la posterior del reconocimiento en rueda; o bien en aquellos casos en que la policía, como consecuencia de las pesquisas efectuadas, o simplemente por conocer ya a aquellos delincuentes que suelen tener un mismo *modus operandi* en hechos similares al/los denunciado/s, sospecha de una o varias personas determinadas y, por ende, aun sin recibir indicación alguna por parte de la víctima o testigo, decide mostrarle diversas imágenes de varias personas con características físicas parecidas para ver si así esta/e refresca su memoria y logra recordar al/los autor/es del hecho concreto.

Resulta fundamental que la diligencia de reconocimiento fotográfico en sede policial se efectue con todas las garantías, de forma absolutamente aséptica y neutra, puesto que en caso contrario las consecuencias para la instrucción de la causa podrían resultar nefastas, tal y como ya se ha advertido. Así, tal y como se dispone en la STS 353/2014, de 8 de mayo (Sala 2ª) FJ 2º "*la diligencia quedaría gravemente viciada si los funcionarios policiales dirigen a los participantes en la identificación cualquier sugerencia, o indicación, por leve o sutil que fuera, acerca de la posibilidad de cualquiera de las identidades de los fotografiados*.", y tal irregularidad podría incluso conllevar la ausencia de valor de la posterior rueda de reconocimiento.

En cuanto a su valor probatorio, si bien es cierto que en caso de que el resultado del reconocimiento fotográfico efectuado en sede policial (con todas las garantías) coincida con el resultado de la diligencia del reconocimiento en rueda posteriormente ratificada en el acto del juicio oral, este suele aportar un plus de fiabilidad a la identificación realizada por la víctima o testigo (habida cuenta de su persistencia y ausencia de contradicción), lo cierto es que la jurisprudencia otorga a los reconocimientos fotográficos policiales un mero valor de actuación previa de investigación (inidóneo e insuficiente para desvirtuar la presunción de inocencia), entendiendo que no pueden considerarse diligencias de reconocimiento de identidad, puesto que tal calificación se reserva para las ruedas de reconocimiento realizadas en fase de instrucción, en sede judicial y con todas las garantías procesales.

Así, la STS 289/2020, de 5 de junio (Sala 2ª) FJ 1º, establece: "*Los reconocimientos fotográficos policiales no constituyen una diligencia de reconocimiento de identidad, sino una actuación previa de investigación, realizada generalmente por la Policía, con la finalidad de orientar adecuadamente las pesquisas encaminadas a la identificación del autor de los hechos. Los reconocimientos de identidad se han de efectuar en ruedas de reconocimiento con la presencia física del sospechoso, que debe estar asistido de letrado, o en el mismo acto del juicio oral (STS 503/2008, de 17 de julio y 1202/2003, de 22 de septiembre).*"

Visto lo expuesto, considero que están claras la problemática y las dificultades que los reconocimientos fotográficos y las ruedas de reconocimiento generan en el proceso de instrucción español, lo que no hace más que incrementar su ineficacia e ineficiencia, por

lo que entiendo que es momento de analizar los posibles beneficios y utilidades que podrían aportar los sistemas de IA de reconocimiento facial, como complemento o, incluso, sustitución de tales diligencias.

Como cuestión previa, no obstante, es importante poner de manifiesto que, tal y como más adelante se desarrollará en más profundidad, las antedichas posibles bonanzas de los mencionados sistemas de IA únicamente serían posibles en aquellos casos en que estos contaran con las máximas garantías de seguridad y pleno respeto a los derechos de los ciudadanos, ya que en caso contrario, sin duda, resultaría "mucho peor el remedio que la enfermedad". Así, y a pesar de los posibles riesgos jurídicos que dichas herramientas de reconocimiento facial pueden entrañar, que como se ha dicho serán analizados posteriormente con más detalle, deben sentarse desde ya las bases sobre las que, a mi entender, estos han de funcionar, a saber: certificación pública de calidad, regulación legal clara, transparencia, explicabilidad, responsabilidad y seguridad.

Dicho esto, y sin perjuicio de lo que luego se pondrá de manifiesto, considero que podrían existir dos opciones para acabar (o, al menos, mitigar) con los problemas que generan las ruedas de reconocimiento en el proceso judicial penal: su sustitución por la utilización de herramientas de IA de reconocimiento facial o su uso como mero complemento.

En primer lugar, sin duda, la forma más drástica y radical de afrontar las dificultades que generan las ruedas de reconocimiento en el proceso penal español sería la de utilizar sistemas de IA de reconocimiento facial en su lugar. Ello, no obstante, en la práctica no siempre resultaría viable, ya que, desafortunadamente, no suele contarse con imágenes (dubitadas) de los presuntos autores de los hechos en el momento de comisión de los mismos o en circunstancias sospechosas que permitan su posterior comparación con las imágenes (indubitadas) que constan o pueden ser introducidas en las bases de datos de tales sistemas.

Y es que en España, como se ha dicho, a diferencia de lo que ocurre en países como China, que sí cuenta con sistemas de vigilancia masiva que implican la presencia de millones de cámaras en la vía pública para controlar cada movimiento de sus ciudadanos, la opción descrita, con carácter general, resultaría impracticable, al menos por el momento. No obstante, procede advertir que incluso en países como el mencionado, tal medida tampoco

tendría siempre el éxito garantizado, habida cuenta de que, como es sabido, muchos actos delictivos se cometen en el ámbito privado o en zonas públicas libres de supervisión.

En virtud de lo expuesto, pues, la aludida sustitución solo resultaría exitosa en aquellos casos concretos en que una cámara (de seguridad o de un teléfono ajeno, entre otros) hubiera captado, por ejemplo, la imagen de una persona cometiendo un delito de robo con violencia en la vía pública y esta resultara posteriormente detenida cerca del lugar de los hechos; así como en aquellos casos en que la imagen de un conductor que atropella a un peatón, se da a la fuga y luego resulta detenido, fuera captada por una cámara de tráfico; o en aquellas ocasiones en que la imagen de un estafador reincidente ya fichado fuera captada por una cámara legalmente colocada en el interior del establecimiento donde ocurren los hechos.

Así, por un lado, imaginemos que un menor de edad es víctima de un delito de homicidio en grado de tentativa, denuncia los hechos a la policía a través de su representante legal, y manifiesta poder reconocer al autor de los hechos (cuya identidad desconoce). Imaginemos también que, afortunadamente, las cámaras de seguridad de la estación de tren donde se produjeron los hechos captaron imágenes de toda la secuencia delictiva, por lo que la policía decide introducirlas en su sistema de IA de reconocimiento facial con el objetivo de averiguar si los datos biométricos del rostro del agresor coinciden con los de alguna de las miles de imágenes de individuos identificados que constan en su base de datos. Y, finalmente, pensemos que el resultado es positivo, puesto que la imagen del rostro del presunto homicida coincide con una de las contenidas en el mencionado sistema, por lo que enseguida se procede a su identificación y se inician los trabajos tendentes a su detención y puesta a disposición judicial, sin necesidad de practicar una ulterior rueda de reconocimiento.

En tal supuesto, como es lógico, de no haberse empleado un sistema de IA de reconocimiento facial por parte de la policía, lo más seguro es que se hubiera acordado el archivo del caso *ab initio* por falta de autor conocido y se hubiera desplegado una ingente cantidad de recursos para poder proceder a la identificación del presunto homicida, sin garantía alguna de éxito. Además, incluso en el caso de que la policía hubiera logrado dar con el sospechoso y se hubiera abierto una investigación judicial frente al mismo, la víctima

menor de edad hubiera tenido que someterse a la práctica de una rueda de reconocimiento, con todas las nocivas consecuencias que ello podría suponer para el joven (y, por supuesto, para los recursos de la policía y del Juzgado) y, asimismo, seguramente, este incluso hubiera tenido que proceder a identificar de nuevo al autor de los hechos en el acto del juicio oral para poder garantizar su condena. No obstante, mediante el uso del sistema de IA de reconocimiento facial, en cambio, la policía desde el primer momento, en cuestión de minutos, hubiera tenido el caso resuelto, haciendo por ende innecesaria la práctica de la diligencia de instrucción de reconocimiento en rueda y, también, bajo mi punto de vista, la identificación del acusado por parte de la víctima en el acto del plenario, siempre y cuando, eso sí, el resultado arrojado por el sistema de reconocimiento facial hubiera podido ser debidamente analizado y sometido a contradicción en dicho acto del juicio oral.

Por otro lado, imaginemos que la víctima de un delito de robo con fuerza en casa habitada llama a la policía en el momento inmediatamente posterior a la comisión de los hechos, le da una descripción del autor y la patrulla policial enviada al lugar identifica en las inmediaciones próximas a la vivienda a un sujeto con características plenamente coincidentes con las expuestas por el testigo y, por ende, procede a su detención, hace la correspondiente reseña fotográfica e introduce la imagen en el sistema de IA de reconocimiento facial. Imaginemos, asimismo, que existen imágenes del momento de la comisión del robo captadas por una cámara de seguridad ubicada en el interior de la vivienda y que la policía, con el fin de comprobar que la persona detenida es la misma que la que consta en dichas grabaciones, procede a introducir en dicho sistema la imagen del rostro que se contiene en estas, que efectivamente confirma que los datos biométricos de uno y otro individuo coinciden.

En tal caso, como es lógico, de no haberse empleado un sistema de IA de reconocimiento facial, la víctima o testigo hubiera tenido que proceder a identificar al autor de los hechos mediante un reconocimiento en rueda, con todo lo que ello conlleva y, sin embargo, mediante el uso de tal herramienta, siempre que, una vez más, sus resultados pudieran ser debidamente sometidos a contradicción en el acto del plenario, la práctica de tal diligencia podría evitarse.

Y es que en tales casos, únicamente habría que llevar a cabo una comparación de imágenes: la dubitada, es decir, la del delincuente captado en el momento de los hechos o en circunstancias sospechosas, y la indubitada, o sea, la que constara ya en la base de datos del sistema o la tomada, por ejemplo, en sede policial al detenido al hacer la reseña, previa introducción en el sistema. Y ello podría hacerse bien con el fin de proceder a la identificación inicial del presunto criminal mediante la comparación de su imagen facial con las de los miles de rostros que constan fichados en la base de datos del sistema de reconocimiento facial empleado, para ver si coincide con alguna, en aquellos casos en que su filiación fuera desconocida; o bien con el fin de proceder a la comprobación de su identidad, mediante la comparación de su imagen facial con la la previamente introducida en la mencionada base de datos, para ver si coinciden, en caso de que ya constara su filiación.

Así, con el resultado de tales operaciones, desde luego, podrían extraerse conclusiones inmediatas sobre la identidad de los autores de los hechos, ahorrando a las víctimas o testigos la práctica de una diligencia tan poco agradable y fiable como es la del reconocimiento en rueda; disminuyendo, además, al mínimo (puesto que un sistema de IA confiable, de calidad, puede llegar a tener niveles de infalibilidad muy elevados) los riesgos de las identificaciones erróneas; reduciendo el despliegue de recursos policiales y judiciales; aumentando la eficacia de la diligencia, que podría practicarse en un corto periodo de tiempo y por cualquier juzgado de España; evitando las dilaciones indebidas, etc.

Visto lo expuesto puede concluirse, pues, que si bien, al menos por el momento, la sustitución genérica de las ruedas de reconocimiento por los sistemas de IA de reconocimiento facial en la práctica no resultaría factible, habrá que ir atendiendo a los avances tecnológicos y jurídicos que vayan incorporándose por las autoridades europeas y de nuestro país para ir valorando sus posibilidades, puesto que, sin duda, valdría la pena avanzar en tal sentido.

En segundo lugar, sin embargo, procede contemplar una opción menos drástica y eficaz pero quizás más viable que la de la sustitución de las ruedas de reconocimiento por el uso de sistemas de reconocimiento facial: la de su complemento.

Y es que, mediante la aplicación de tal solución, en mi opinión, si bien no se erradicarían todas las dificultades que actualmente acechan en la práctica de las diligencias de reconocimiento en rueda, lo que sí se lograría es, por un lado, conseguir más indicios de criminalidad frente a una persona investigada o acusada de cometer un delito; por otro lado, asistir a los órganos judiciales en su labor de valoración de los resultados de las ruedas de reconocimiento; y, finalmente, dotar a los investigados o acusados de una mayor seguridad jurídica y garantía de respeto a sus derechos fundamentales. Me explico.

Por una parte, imaginemos que se ha cometido un delito de agresión sexual, la víctima llama a la policía instantes más tarde, le explica lo sucedido y manifiesta que en caso de volver a ver al autor de los hechos no podría reconocerlo, puesto que la zona donde estos se cometieron estaba poco iluminada, era de noche y fue atacada por la espalda. Imaginemos, asimismo, que lo único que la víctima puede asegurar es que el agresor le arrebató una chaqueta roja que llevaba puesta y que tenía acento extranjero. Sigamos imaginando y, pensemos, que un vecino de la zona, tras ser interrogado por la policía, que acude inmediatamente al lugar de los hechos, asegura que instantes antes se ha cruzado con un varón que le ha pedido tabaco con acento extranjero y llevaba colgada una chaqueta roja en el hombro, si bien manifiesta que le resultaría complicado reconocerlo porque el contacto visual ha sido muy rápido y escaso. Y, finalmente, imaginemos que la policía, con tales datos, busca por las proximidades y encuentra a un hombre sin aparente rumbo, le da el alto y, a pesar de que no lleva chaqueta roja alguna, responde con acento extranjero, de forma nerviosa, por lo que se procede a su detención y se le toma una fotografía que luego es introducida en un sistema de IA de reconocimiento facial para su comparación con todas aquellas imágenes de rostros de personas que constan en una base de datos de delincuentes sexuales habituales, resultando coincidente con una de ellas. Tras ello, imaginemos que la policía realiza sus pesquisas, averigua que tal individuo fue condenado en el pasado por dos delitos de agresión sexual y descubre que en las dos ocasiones atacó a sus víctimas por la espalda, de noche, en lugares oscuros y cerca de donde ocurrieron los hechos objeto de la actual investigación, lo cual, desde luego, si bien nunca podría resultar una prueba de cargo única y bastante para desvirtuar la presunción de inocencia de la persona investigada, sí que podría ser valorado como un indicio más tanto por el juez de instrucción como por el órgano enjuiciador. Y, finalmente, pensemos que la persona detenida es sometida a un reconocimiento en rueda por parte del vecino testigo y este no logra identificarlo y, por su

parte, la víctima procede al reconocimiento de la voz del agresor en el acto del plenario, pero afirma tener dudas.

Así, en el caso expuesto, de no haberse utilizado un sistema de IA de reconocimiento facial lo más probable es que, ante la ausencia de indicios racionales frente a la persona detenida (más allá de su acento extranjero, su presencia en las proximidades del lugar del crimen poco después de los hechos y el reconocimiento dudoso de la voz realizado por la víctima), la causa hubiera podido quedar sobreseída ya en fase de instrucción por desonocimiento de autor o, en su caso, el acusado hubiera podido ser absuelto en aplicación del principio *in dubio pro reo*. Sin embargo, con el uso del resultado del análisis comparativo del sistema de IA, lo que se consigue es aportar nuevos indicios muy valiosos para la investigación de la causa y su posterior enjuiciamiento, puesto que se cuenta con varios indicios compatibles a valorar en conjunto: la declaración de la víctima y el reconocimiento de voz (aunque dudoso), la declaración testifical del vecino y de los policías que procedieron a la detención del sospechoso, el resultado del sistema de IA de reconocimiento facial y el posterior informe policial que evidenciaría el pasado delictivo (en que constaría el mismo *modus operandi* que el de los hechos investigados) de la persona acusada, lo cual sin duda -especialmente en caso de que el acusado no diera una versión exculpatoria lógica y creíble-, a pesar del fracaso de la rueda de reconocimiento efectuada por el testigo, podría llegar a conllevar la condena del acusado.

Asimismo, resulta interesante pensar en un supuesto en que, por ejemplo, una víctima de un delito de amenazas con instrumento peligroso acuda a la policía a denunciar los hechos y otorgue una descripción clara y precisa del autor (cuya identidad desconoce), permitiendo así la realización de un retrato robot de alta calidad. Imaginemos que, en tal caso, la policía procediera, mediante el uso de un sistema de IA de reconocimiento facial, a comparar tal imagen robot con las imágenes legalmente contenidas en su base de datos y resultara que el rostro plasmado en tal retrato fuera coincidente con el de alguno de los individuos ya "fichados". Imaginemos, asimismo, que tras la identificación del posible sospechoso a través del uso de tal sistema, por un lado, la policía sometiera a la víctima a un reconocimiento fotográfico, y por otro lado, el juez de instrucción acordara, de forma adicional, la práctica de un reconocimiento en rueda, lo cual, sin duda, podría resultar clave

para la continuación o el archivo de la causa (y, en su caso, el pronunciamiento absolutorio o condenatorio de la sentencia), en función del éxito o el fracaso de tal diligencia.

Así, en el caso expuesto, de no haber sido empleado un sistema de IA de reconocimiento facial, el caso hubiera quedado archivado desde el principio por desconocimiento de autor (salvo en el improbable supuesto de que los policías hubieran identificado al sospechoso del retrato robot por conocerlo de actuaciones anteriores y hubieran podido practicar un reconocimiento fotográfico del mismo con la víctima, que posteriormente hubiera sido sometida a una rueda de reconocimiento) y, sin embargo, mediante la utilización de tal herramienta tecnológica, se da la posibilidad de que la investigación avance, se realice en sede judicial una rueda de reconocimiento y el caso acabe, incluso, con una sentencia condenatoria tras su ratificación y contradicción en el acto del plenario.

No obstante, en el supuesto comentado, en mi opinión bastaría con el mero uso del sistema de reconocimiento facial y la posterior identificación del acusado por la víctima y/o los testigos en el acto del juicio oral -siendo que la jurisprudencia admite su valor identificativo, tal y como se desprende, entre otras, de la STS 786/2017, de 30 de noviembre (Sala 2ª)-, básicamente para evitar a esta/os tener que pasar por la práctica de una rueda de reconocimiento y lograr así un ahorro de recursos policiales y judiciales, pero entiendo que la opción expuesta resulta menos drástica y puede erigirse como perfecto término medio para aquellos más garantistas o conservadores.

Por otra parte, imaginemos que en un caso de comisión de un delito de lesiones agravadas, una víctima identifica al presunto autor de los hechos mediante un reconocimiento fotográfico policial, manifestando, no obstante, tener serias dudas al respecto. Imaginemos también que tal persona identificada, a raíz del reconocimiento realizado por la víctima en sede policial resulta investigada y, por ende, se acuerda por el juez de instrucción la práctica de una rueda de reconocimiento en que la testigo, de edad muy avanzada, identifica al presunto autor pero manifiesta que solo está segura de su identidad al 40-50%, asegurando que no recuerda cómo iba vestido en el momento de los hechos, puesto que solo se fijó en el rostro. No obstante, pensemos también que existen unas imágenes captadas por una cámara de seguridad de la vía pública colocada a escasos metros del lugar de comisión de los hechos que tan solo unos instantes después de los mismos muestran a un individuo

caminando solo a paso ligero, lo que levanta las sospechas de la policía que, no obstante, resulta incapaz de reconocerlo y asegurar que se trata de la misma persona que se halla investigada, puesto que lleva gorra y gafas. Imaginemos, asimismo, que se acuerda la práctica de un informe pericial antropomórfico que tampoco arroja luz a la investigación puesto que no resulta concluyente. Y, finalmente, imaginemos que, de forma paralela, se emplea por la policía un potente sistema de IA de reconocimiento facial que, a pesar de la gorra y las gafas, es capaz de identificar a la persona que se hallaba en la vía pública a escasos metros y minutos del lugar de los hechos con el individuo reconocido (con dudas) por la víctima en el reconocimiento en rueda.

Así, sin duda, el uso de una herramienta de IA de reconocimiento facial en tal caso decantaría la balanza hacia la continuación del procedimiento y, seguramente, hacia una sentencia condenatoria de la persona identificada (siempre tras la debida contradicción llevada a cabo en el acto del plenario), puesto que vendría a reforzar y confirmar lo que la víctima había sido incapaz de asegurar al 100% en la diligencia de reconocimiento en rueda, debido a su vulnerabilidad por razón de edad.

No obstante, y siguiendo con la misma línea de opinión anteriormente expuesta, entiendo que en casos como este último bastaría con la utilización de sistemas de IA de reconocimiento facial en sede policial y judicial (incluso considero que con el judicial sería suficiente) y la posterior identificación del autor de los hechos por parte de la/s víctima/s o testigo/s en el acto del plenario, justamente para evitar las ya conocidas perniciosas consecuencias de la práctica de la rueda de reconocimiento, pero soy consciente de que se trata de una opción un tanto disruptiva.

En relación con los casos anteriormente expuestos, procede poner de manifiesto que, sin duda, el auxilio en la tarea de valoración de pruebas para el órgano judicial alcanzaría su máximo nivel en aquellos casos en que, existiendo imágenes del autor de los hechos en el momento de su comisión, este fuere identificado tanto por las víctimas o testigos mediante reconocimiento fotográfico policial y posterior reconocimiento en rueda en sede judicial, como por el sistema de IA de reconocimiento facial empleado por la policía y por el juzgado, ya que ello cerraría la instrucción de forma casi inatacable y dejaría un fácil camino al órgano enjuiciador para el acto del plenario, donde tan solo tendrían que

someterse los más que concluyentes resultados a contradicción, con un más que probable posterior dictado de sentencia condenatoria.

Finalmente, procede poner de relieve que el uso de sistemas de IA de reconocimiento facial como complemento a las ruedas de reconocimiento ayudaría también, tal y como ya se ha advertido con anterioridad, a dotar de mayores garantías procesales a los investigados o acusados, puesto que al igual que podrían servir para reforzar o confirmar los resultados de tales diligencias de investigación, podrían también servir para hacerlos tambalear y restarles valor en caso de arrojar conclusiones dispares.

Así, imaginemos que en un caso de delito por lesiones ocasionadas en una pelea callejera grabada por un sistema de videovigilancia, la víctima procede a reconocer, tanto fotográficamente como en rueda, al presunto autor de los hechos. No obstante, imaginemos que si bien el sistema de reconocimiento facial empleado por la policía coincide con el veredicto del perjudicado, el sistema de reconocimiento facial utilizado por el Juzgado discrepa de este y no identifica como autor a la persona inicialmente investigada. Y es que en un caso así, la polémica en el acto del plenario (e incluso en fase de instrucción) estaría servida, habida cuenta de que el sistema tecnológico empleado por la Administración de Justicia habría hecho tambalear, sin duda, las conclusiones extraídas de la diligencia de investigación de reconocimiento en rueda efectuada por la víctima y de las diligencias policiales, lo cual podría conllevar, en caso de no existir otras pruebas concluyentes, la aplicación del principio *in dubio pro reo*, lo que de otra forma, sin embargo, no hubiera resultado posible.

Analizado todo lo anterior, pues, bajo mi punto de vista, desde el mismo momento en que tengamos a nuestra disposición sistemas de IA de reconocimiento facial confiables, de calidad, (en los términos expuestos anteriormente, que posteriormente se analizarán con más profundidad), estos deberían ser empleados en fase de instrucción, sin duda, tanto para sustituir a las ruedas de reconocimiento cuando ello resulte posible, como para complementarlas en caso de que solo exista tal posibilidad (con posterior identificación de los presuntos criminales por parte de las víctimas o testigos en el acto del plenario cuando resulte necesario). Y es que entiendo que, aunque cueste dar pasos en tal sentido, por entenderse arriesgado, hacerlo de forma determinante y regulada es la única forma de no

quedarnos anclados en sistemas pretéritos que no resultan eficientes y de avanzar hacia un proceso penal tecnológico de calidad, con todas las garantías.

No obstante, no hay que perder de vista que ello, en la actualidad, es todavía una utopía, incluso en el caso de los sistemas de reconocimiento facial más potentes, puesto que estos podrían dar problemas de identificación, ya que, tal y como ya se ha puesto de manifiesto con anterioridad, tales herramientas, al menos por el momento, aun son muy inmaduras y tienen grandes limitaciones.

a.2.2.) Identificación de los presuntos delincuentes mediante la comparación de las imágenes de los hechos, captadas por un dispositivo fotográfico o de vídeo, con las imágenes contenidas en las bases de datos de los sistemas de IA de reconocimiento facial en aquellos casos en que no haya testigos

No es infrecuente, en la actualidad, que tras la comisión de un crimen sin testigos, el Ministerio Fiscal, la acusación particular o la defensa soliciten al juez de instrucción que acuerde oficiar a un centro penitenciario, a un Ayuntamiento, a un local de ocio o a un comercio para que remitan las imágenes captadas por sus cámaras de seguridad con el fin de poder hallar indicios de la autoría de los hechos.

En tales casos, no obstante, al no contarse con la identidad del autor que aparece en las imágenes, son los propios agentes de policía los que las visualizan y tratan de averiguar de quién se trata, lo cual no siempre tiene resultados de éxito, habida cuenta de la baja probabilidad que existe de que un delincuente sea conocido justamente por los agentes de policía que examinan las mismas (aunque, evidentemente, en aquellos casos de personas reincidentes en una misma zona y un mismo periodo de tiempo, suele funcionar).

Ante ello, y para evitar las enormes limitaciones de tan tradicional (y, por desgracia, poco efectivo) método, entiendo que la utilización de sistemas de IA de reonocimiento facial podría implicar una revolución sin precedentes.

Imaginemos, por ejemplo, que una cámara de tráfico capta a un infractor varón conduciendo un vehículo a 200km/h por una carretera nacional cuya limitación de velocidad son 100 km/h. Imaginemos también que, llegada tal *notitia criminis* a la policía o al Juzgado, se introduce la matrícula del vehículo en la base de datos policial o en el

sistema de Punto Neutro judicial y la información que se obtiene es que la titular de tal vehículo es una mujer, que es llamada a declarar como testigo y manifiesta que no está segura de si el día de los hechos cogió el coche su hermano o su cuñado, puesto que ambos tenían llaves. Pensemos, asimismo, que se cita a declarar como investigados tanto al hermano como al cuñado, ambos niegan haber conducido el vehículo el día de los hechos y la policía, por las imágenes que constan en las grabaciones, no se ve capaz de distinguir con claridad si el conductor era uno u otro, ya que ambos tienen rasgos físicos similares. Imaginemos, no obstante, que se realiza una fotografía de frente a cada investigado y se introduce en el correspondiente sistema de reconocimiento facial para comparar con las imágenes captadas por la cámara de tráfico, lo que permite, en cuestión de segundos, identificar al conductor, que resulta ser el hermano de la titular del coche.

Así, en tal supuesto, de no haber sido empleado un sistema de IA de reconocimiento facial, posiblemente el caso hubiera quedado archivado por desconocimiento de autor o los acusados hubieran quedado absueltos en sentencia en aplicación del principio *in dubio pro reo* y, sin embargo, mediante el uso de tal herramienta, se consigue la identificación del delincuente de forma efectiva en un corto lapso de tiempo, lo que permite que este resulte acusado y sea posteriormente juzgado en el acto del plenario, siempre con la debida posibilidad de contradicción de los resultados que arrojara el sistema de IA.

Imaginemos, asimismo, que una cámara de videovigilancia legalmente instalada en una tienda de telefonía capta las imágenes de un individuo que, tras el cierre, fuerza la persiana, entra y rompe la caja registradora, de donde sustrae la recaudación del día. Imaginemos, asimismo, que el propietario de la tienda lleva tales imágenes a la policía, que procede a su visualización y advierte que el autor de los hechos es un varón que portaba una capucha y resulta inidentificable por no existir indicio alguno que pueda llevar a su filiación. Pensemos también que al cabo de dos días una trabajadora de dicho comercio observa, cerca de la tienda, a un hombre que lleva a cabo comportamientos extraños, tales como mirar con mucho detalle el escaparate, hacer fotografías, preguntar por los precios y el *stock* de los modelos de teléfonos de mayor valor etc, por lo que llama a la policía, que se persona en el lugar y, ante los últimos acontecimientos, pregunta a tal individuo qué hace por allí, a lo que este responde con evasivas, por lo que es trasladado a comisaría para ser identificado. Imaginemos que, en ese momento, los agentes de policía visualizan las

imágenes aportadas por el propietario de la tienda, pero al ver que el individuo que consta en ellas llevaba capucha, no se ven capaces de afirmar con seguridad que las identidades coinciden, por lo que toman una fotografía al sujeto identificado, que es introducida en un potente sistema de IA de reconocimiento facial, junto con las mencionadas imágenes captadas por las cámaras de videovigilancia, y ello lleva a detectar en un cortísimo plazo de tiempo que los datos biométricos de los rostros que aparecen en ambas imágenes coinciden, por lo que se procede a su detención y posterior investigación.

En adición, imaginemos que la policía, durante el último mes, hubiera recibido denuncias procedentes de hasta quince establecimientos de telefonía de diversas localidades asegurando haber sufrido robos con fuerza captados por las imágenes de sus cámaras de seguridad, permaneciendo, sin embargo, los autores sin identificar. Así, pensemos que la policía, al observar cierta similitud entre los distintos hechos (horario, *modus operandi*, tipo de tienda, etc), decidiera introducir la imagen de la persona identificada (y, posteriormente, detenida) en el correspondiente sistema de IA de reconocimiento facial junto con las imágenes obtenidas por los sistemas de videovigilancia de los diversos establecimientos y este arrojara un resultado de coincidencia biométrica positiva en el 100% de los casos. Ello, sin duda, lo que implicaría es la imputación a tal sujeto no solo del robo con fuerza del último establecimiento, sino también de los otros quince que permanecían sin resolver.

Así, en tal supuesto, de no haberse empleado un efectivo sistema de IA de reconocimiento facial, la condena de la persona identificada días después de los hechos en las proximidades del último comercio donde se produjo el robo hubiera sido casi imposible y, sin embargo, mediante el uso de tal herramienta, esta quedaría prácticamente asegurada. Por no mencionar la condena por el resto de delitos perpetrados, que hubieran permanecido en archivo por falta de autor hasta quizás acabar archivados por prescripción.

Pensemos, también, en aquellos casos en que un ciudadano denuncia un robo con fuerza en interior de vehículo sin contar con información alguna sobre el presunto autor de los hechos. Imaginemos, asimismo, que al cabo de los años aparecen unos vídeos de los hechos denunciados en el teléfono móvil de un delincuente común que estaba siendo investigado por otros delitos. Imaginemos, también, que ante tal hallazgo casual, el juez de instrucción autoriza el volcado y análisis de tales imágenes y procede a emplear un sistema de

reconocimiento facial para averiguar si el rostro del individuo que figura en los vídeos de los hechos contenidos en el teléfono interceptado coincide con el de alguna de las imágenes que constan en su base de datos, arrojando ello un resultado positivo que permitiera dar con la identidad del delincuente.

Así, en tal caso, de no haberse empleado un sistema de IA de reconocimiento facial, el caso hubiera quedado, con toda seguridad, sobreseído por desconocimiento de su autor, y sin embargo, de esta forma, con alta probabilidad (siempre con garantía de contradicción en el plenario), el asunto hubiera acabado en condena.

Y, finalmente, pensemos en aquellos casos en que si bien existien grabaciones del autor de, por ejemplo, un delito de daños en mobiliario público en el momento de su comisión, se desconoce su identidad, lo cual con toda probabilidad conduciría a un archivo inmediato de la causa por desconocimiento de autor. Imaginemos, no obstante, que la policía cuenta con un sistema de IA de reconocimiento facial e introduce en el mismo la imagen del rostro del delincuente (dubitada) para proceder a su comparación con las otras miles de imágenes faciales que constan en su base de datos (indubitadas), lo que permite llevar a cabo su filiación con éxito.

Así, en tal caso, el uso de un sistema de reconocimiento facial ayudaría a identificar al posible autor de los hechos y habría dado la posibilidad de que, al menos, este hubiera sido sometido a juicio en el acto del plenario, lo cual hubiera resultado impensable sin la utilización de tal tecnología.

Y es que, sin duda, vemos que el uso de sistemas de reconocimiento facial podría ayudar a resolver miles de asuntos que, en caso contrario, resultarían archivados y posteriormente prescritos.

De hecho, en mi opinión, en la práctica totalidad de los casos expuestos, es muy probable que el Letrado del investigado, al advertir el concluyente resultado arrojado por el sistema de reconocimiento facial (siempre en caso de que se reputara confiable), aconsejara a su cliente que reconociera los hechos, lo que implicaría una transformación del procedimiento de Diligencias Previas en Diligencias Urgentes y acabaría con una más que probable

sentencia de conformidad, con el ahorro de recursos materiales y personales que ello supone para la Administración de Justicia.

a.2.3) Identificación de los autores de un delito mediante la captación de imágenes a través de gafas inteligentes empleadas por la policía

En algunas ocasiones, los agentes de policía se hallan realizando tareas de vigilancia ordinaria cuando se ven sorprendidos por la comisión de un delito o, en otras, son comisionados para acudir rápidamente al lugar donde este se está perpetrando. Como consecuencia de sus actuaciones, no obstante, y dependiendo de múltiples circunstancias, en algunas ocasiones los agentes consiguen detener e identificar a los presuntos criminales, pero en otras, estos logran escapar sin ni siquiera resultar identificados, lo que implica una posterior (y, por lo general, ardua) tarea de investigación que no siempre finaliza con éxito.

En relación con ello, procede hacer mención de la existencia de una herramienta de IA que, sin duda, podría resultar de gran utilidad para la fuerza pública en tales escenarios y, por consiguiente, podría aportar grandes beneficios a la posterior instrucción judicial de las causas: las gafas inteligentes que llevan incorporados sistemas de IA de reconocimiento facial.

Actualmente, ya existen dispositivos con forma de gafas (o *body cameras*) diseñados para ser usados por los agentes de policía con la finalidad de captar imágenes instantáneas, en tiempo real, de aquellas personas con las que esta se cruza y considera de interés (por múltiples motivos) y así, de forma inmediata, proceder a su introducción en una base de datos para someterlas a comparación con las otras miles o millones de imágenes que ya constan en el sistema, a través de un algoritmo, con fines identificativos.

Así, en países como China, por ejemplo, el uso de las mencionadas gafas inteligentes por parte de los agentes de policía ha resultado ampliamente extendido desde su introducción en el año 2018.

Y es que en 2017 la empresa Xloong, con sede en Pequín, diseñó unas gafas inteligentes de realidad aumentada para uso policial vinculadas con la extraordinariamente extensa base de datos nacional, siendo su principal finalidad que los agentes de policía pudieran acceder a información en tiempo real sobre el rostro, la tarjeta de identificación y la placa del

vehículo de las personas con las que se cruzaran (especialmente si eran sospechosas de haber cometido algún delito).[586]

De tal modo, los agentes de policía chinos que llevan las mencionadas gafas inteligentes pueden identificar a presuntos delincuentes de forma prácticamente instantánea y entre multitud de personas, simplemente haciéndoles una fotografía a través de las mismas. Y es que tales dispositivos, de forma automática introducen en la mencionada base de datos la imagen captada para que un sistema de IA de reconocimiento facial pueda compararla con los otros millones de imágenes en ella contenidas, aportando así al agente, en caso de éxito, no solo la filiación de la persona fotografiada sino también otros datos de interés, tales como su dirección postal.[587]

[588]

Así, por ejemplo, imaginemos que una organización criminal integrada por cuatro miembros ("teloneros") se dedica a cometer, por todo el territorio nacional, robos de mercancía de camiones mientras estos se hallan aparcados por la noche, habiendo llegado a sustraer productos por importes millonarios. Imaginemos, también, que un día la policía recibe la llamada de un conductor de camión que manifiesta que acaba de ser víctima de un robo de mercancía y que le parece estar escuchando cómo se está cometiendo otro robo en un camión cercano, por lo que la policía de inmediato manda una patrulla al lugar de los hechos. Pensemos que, al llegar, los agentes observan a dos personas apoyadas en un coche, en actitud vigilante, y a dos individuos más saliendo de un camión cargados con productos

[586] Véase más en Yingzhi, 2019.
[587] Véase BBC News, 2018.
[588] Yang, 2019.

alimentarios, por lo que se dirigen hacia ellos para proceder a su detención. No obstante, imaginemos que estos, que están perfectamente organizados, consiguen meterse rápidamente en el vehículo y huir, dando lugar a una peligrosa persecución policial que finaliza sin éxito, lo que conlleva su impunidad, siendo que no se cuenta con pista alguna sobre su identidad.

En tal caso, sin embargo, imaginemos que los agentes hubieran llevado unas gafas inteligentes y en el mismo momento de llegada al lugar de los hechos hubieran podido captar, de forma instantánea, imágenes de todos los sospechosos, de forma que en cuestión de minutos hubieran dado con sus identidades y demás datos personales. Así, a pesar de la huida, la policía hubiera podido contar con una información valiosísima para iniciar una investigación que, con toda probabilidad, hubiera acabado judicializándose y, en caso de localizar a los sospechosos, hubiera terminado en la condena de los mismos, no solo por los últimos hechos sino por todo el resto, en caso de haber podido comparar las imágenes de sus rostros con las eventualmente tomadas por las cámaras de vigilancia ubicadas en los distintos lugares de comisión de los hechos.

a.2.4) Identificación, detención y puesta a disposición policial o judicial, mediante gafas inteligentes dotadas de sistemas de IA de reconocimiento facial, de las personas investigadas sobre las que pesan órdenes de detención policial o de busca y captura judicial

En ocasiones, si bien los presuntos delincuentes de los hechos se hallan identificados, no es posible determinar su paradero para someterlos a las diligencias de investigación policial o judicial que restan pendientes o, incluso, para citarlos para la celebración del acto del juicio oral o para notificarles la correspondiente sentencia. Ante tales circunstancias, los agentes de policía emiten órdenes de detención, si la investigación se halla en fase policial, y los jueces emiten órdenes de busca y captura (requisitorias judiciales), en virtud de lo dispuesto en los artículos 512 y ss de la LECrim, si el caso se halla judicializado.

Tales medidas, no obstante, no siempre resultan suficientes y efectivas, ya que si bien en algunas ocasiones la policía acaba dando con los individuos buscados, en muchas otras ello no resulta posible o se demora durante años, lo que se traduce en causas judiciales eternas, prescripción de las mismas y posteriores archivos, con la consiguiente impunidad que

conlleva para los delincuentes. Y es que este es uno de los grandes problemas a los que se enfrentan los investigadores de delitos de todos los países del mundo, puesto que, salvo que cuenten con indicios o pistas muy claras que les ayuden a localizar a la/s persona/s huída/s, lo cierto es que, por lo general, carecen de información relevante a efectos de llevar a cabo dichos hallazgos y, por ende, su éxito queda sometido a la "suerte" de que estas sean paradas, de forma aleatoria, en un control policial, acudan a un establecimiento hotelero y faciliten su DNI, reincidan en la comisión delictiva y sean detenidos por otra causa, etc.

Mediante el uso de gafas inteligentes, no obstante, un agente de policía en ruta de vigilancia ordinaria tendría la posibilidad de ir identificando, en cuestión de segundos, en tiempo real, a aquellas personas con las que se fuera cruzando y estuvieran etiquetadas como "buscadas" en la base de datos vinculada al sistema de IA, puesto que la cámara de las lentes, dotada de tecnología de reconocimiento facial, sería capaz de detectar los datos biométricos de estas y compararlos con los de las imágenes previamente introducidas en tal sistema, haciendo saltar la alarma en caso de coincidencia.

Así, en el ejemplo anteriormente expuesto, en caso de que los agentes de cualquier punto de España o Europa hubieran podido identificar, a través de las imágenes captadas por sus gafas inteligentes, a los "teloneros" posteriormente huidos, la tarea de detención de los mismos y ulterior puesta a disposición judicial podría haberse visto muy facilitada. Y es que así, entre multitudes de personas, cualquier agente de policía del país podría haber detectado los datos biométricos de los individuos introducidos como sospechosos del robo de material de camiones en su base de datos, lo que hubiera posibilitado su localización y posterior detención. De tal forma, en dicho supuesto, con el mero cruce por la vía pública de los sospechosos de los hechos con un agente de policía, este hubiera podido advertir su presencia y haber procedido a su detención, lo cual hubiera resultado imposible en caso de no emplearse tal tecnología.

a.2.5) Hallazgo de personas desaparecidas a través de las imágenes de las cámaras de vigilancia y/o las gafas inteligentes de la policía

Sin duda, considero que los sistemas de reconocimiento facial podrían resultar de gran utilidad para resolver los (casi siempre complejos) casos de personas desaparecidas, ya que al introducir imágenes de estas en una base de datos, las cámaras de vigilancia podrían

alertar de forma inmediata a la policía en tiempo real cada vez que hallaran coincidencia biométrica con los viandantes, lo cual sería de enorme ayuda.

Así, por ejemplo, la policía de la India, que emplea técnicas de reconocimiento facial para fines de investigación, en abril de 2018 localizó, en tan solo cuatro días, a casi tres mil niños desaparecidos que se habían perdido o habían sido secuestrados. Y es que con tal fin, el Ministerio de Desarrollo de la Mujer y del Menor del mencionado país creó una base de datos nacional denominada "*TrackChild*", donde las autoridades policiales introdujeron miles de fotografías de niños desaparecidos que fueron posteriormente sometidas a análisis por un sistema de reconocimiento facial introducido en las cámaras de vigilancia de Nueva Delhi, que logró identificar a dos mil novecientos treinta niños en el período comprendido entre el 6 y el 10 de abril y alertó de ello a la policía, que de forma inmediata procedió a reunirlos con sus familias.[589]

Y es que es de sobra conocido el enorme sentimiento de angustia y desesperación que suele envolver los terribles casos de personas desaparecidas, habida cuenta de la incertidumbre que estos entrañan y de la, por lo general, enorme complejidad de las investigaciones, que en muchas ocasiones acaban cayendo en saco roto debido a la falta de indicios o de pruebas que permitan llevar a localizar a las víctimas, con el desasosiego y frustración que ello conlleva no solo para las familias sino también para las autoridades policiales, fiscales y judiciales encargadas del asunto.

Así, históricamente, en muchas ocasiones hemos escuchado aquello de que buscar a una persona desaparecida es como tratar de "encontrar una aguja en un pajar", y muchas veces tal símil, por desgracia, resulta cierto. No obstante, con el uso masivo de tecnologías de reconocimiento facial, las tareas de búsqueda de personas desaparecidas devendrían infinitamente más sencillas, rápidas y efectivas, habida cuenta de que, tal y como se ha expuesto en el ejemplo de la India, bastaría con introducir sus fotografías en una base de datos que conectara con el sistema de reconocimiento facial empleado por las cámaras de seguridad de la vía pública y, en su caso, con las gafas inteligentes empleadas por la policía, para que, de detectarse coincidencias biométricas, se emitiera una alerta que permitiera a

[589] Véase India Today Web Desk, 2018.

las fuerzas del orden actuar de forma inmediata y, con casi total seguridad, localizar a la persona perdida y verificar si se trataba de un caso de desaparición voluntaria o forzosa.

a.2.6.) Identificación inmediata de presuntos autores de delitos de quebrantamiento de medida cautelar o de condena

Tanto en fase de instrucción como en fase de ejecución de sentencia existen dos tipos delictivos que son significativamente frecuentes: el quebrantamiento de medida cautelar, en el primer caso, y el delito de quebrantamiento de condena, en el segundo.

Tales delitos, si bien son de aparente fácil instrucción, en ocasiones conllevan dificultades que implican un despliegue de recursos materiales y personales desproporcionado. Y es que, salvo en aquellos casos en que la policía sorprende al delincuente *in fraganti*, este tipo de delitos requiere la declaración testifical de una o más personas y la práctica de otras diligencias (tales como, por ejemplo, ruedas de reconocimiento), lo que implica que una instrucción que podría resultar prácticamente inmediata o muy corta, se alargue a veces innecesariamente.

Así, por un lado, imaginemos que en un caso de violencia de género se impone por parte del juez de guardia una orden de alejamiento, con prohibición de entrada del marido en la población donde reside la mujer. Imaginemos, asimismo, que un día la esposa sale a comprar el pan y se encuentra al todavía esposo sentado en un banco mirándola de modo desafiante, a pesar de la mencionada prohibición, por lo que esta huye corriendo y llama a la policía, que cuando llega al lugar de los hechos ya no encuentra al infractor. Pensemos, asimismo, que posteriormente la policía acude al domicilio del mencionado varón y no lo encuentra, no pudiendo detenerlo hasta dos días después.

En tal caso, como es evidente, únicamente se contaría con el testimonio de la mujer, por lo que esta, para que el caso pudiera prosperar, se vería obligada a acudir en primer lugar a interponer una denuncia a la comisaría de policía, luego asistir al Juzgado de Instrucción a prestar declaración como testigo y, finalmente, hacer presencia en el Juzgado de lo Penal para ratificar su declaración en el acto del plenario, con todos los efectos psicológicos nocivos que ello podría suponerle, además de la gran cantidad de recursos personales y materiales que deberían emplearse (funcionarios de policía que le recogieran la denuncia,

Letrado de la acusación -particular o de oficio-, Juez que le tomara declaración en presencia del Ministerio Fiscal, de dicho Letrado y del de la defensa, citaciones varias, etc).

No obstante, en caso de que se instalaran cámaras de videovigilancia en todos los puntos de acceso y de salida de las distintas poblaciones, que fueran directamente conectadas a una base de datos que contuviera las imágenes de aquellas personas sometidas a prohibición de entrada en las mismas y emplearan sistemas de IA de reconocimiento facial, podrían detectarse en tiempo real, en cuestión de segundos, los quebrantamientos de medida cautelar. Y es que la idea es que tales sistemas, al hallar coincidencia entre los datos biométricos faciales de la persona que accediera dentro del perímetro de la población y los que constaran en alguna de las imágenes contenidas en la base de datos, procedieran a remitir, de forma automática, una alarma a la policía que, de modo instantáneo, tendría conocimiento de la comisión del mencionado delito y podría acudir de forma inmedia al lugar de los hechos para evitar males mayores.

Así, con el uso de sistemas de IA de reconocimiento facial, la instrucción de la causa por quebrantamiento de medida cautelar podría realizarse en apenas un día, siendo que la principal prueba radicaría en la detección por parte del sistema de IA de la entrada del individuo sobre el que pesaba la medida cautelar de prohibición de entrada en la población (siempre que los resultados pudieran ser objeto de contradicción por la defensa) y, a parte de eso, únicamente procedería, por un lado, exhortar al Juzgado que dictó la mencionada medida cautelar para que remitiera al Juzgado de Guardia testimonio de la notificación y requerimiento de cumplimiento del Auto que la acordó; y, por otro lado, tomar declaración en calidad de investigado al presunto autor de los hechos, no resultando necesario, en mi opinión, que la mujer tuviera intervención alguna en el procedimiento (o que esta fuera mínima, de simple ratificación) -salvo en aquellos casos en que además del quebrantamiento de medida cautelar se hubiera cometido algún otro delito no captado por las cámaras-, lo cual, no solo ahorraría recursos personales y materiales, sino que también evitaría a esta la nociva carga emocional de tener que pasar por un nuevo procedimiento judicial frente a su agresor. Además, bajo mi punto de vista, la existencia de tan concluyentes resultados por parte de los sistemas de IA de reconocimiento facial implicaría que la mayoría de las personas investigadas reconociera ya de entrada los hechos, lo que conllevaría la incoación de un procedimiento de Diligencias Urgentes (o, en su caso, la

transformación) que en muchas ocasiones finalizaría con sentencia de conformidad, con el ahorro de recursos que ello comporta para la Administración de Justicia.

Por otro lado, imaginemos que una persona resultara condenada en firme por un delito continuado de hurto cometido en el metro de una ciudad y, parte de la condena, implicara la prohibición de entrada en tal medio de transporte público durante 2 años. Imaginemos, también, que esta persona, a pesar de la condena, decidiera entrar en el metro y reincidir, cometiendo más de 15 hurtos sin ser interceptada por la policía.

En tal supuesto, no obstante, en caso de que existieran cámaras dotadas de sistemas de reconocimiento facial en los accesos de las diversas estaciones de metro de la ciudad, la indebida actuación de la infractora podría haberse gestionado de una forma muchísimo más eficiente y exitosa. Y es que, en tal supuesto, desde el primer momento en que el dispositivo hubiera captado la imagen de la mencionada delincuente entrando en el metro en contra de la prohibición impuesta, hubiera remitido una alerta automática a la policía (e incluso, para ser más efectivo, al personal de seguridad del metro), lo que hubiera implicado no solo el inicio de una sencilla y rapídisima instrucción del delito de quebrantamiento de condena cometido, sino también la más que probable evitación de los nuevos delitos de hurto perpetrados.

En relación con lo anterior, procede hacer especial referencia a la novedosa y eficaz herramienta de IA que la cadena de supermercados MERCADONA instaló en varios de sus establecimientos para detectar justamente la comisión de delitos de quebrantamiento de medida cautelar y de condena.

Así, tal y como me expuso Daniel Larios Caparrós, responsable de seguridad de tal cadena de supermercados en Barcelona Sur y Tarragona, a lo largo de varias conversaciones mantenidas, MERCADONA adquirió de la empresa israelí AnyVision un *software* de reconocimiento facial que permitía la identificación en 0,3 segundos, a través de las cámaras ubicadas en el interior de sus supermercados, de aquellas personas que accedían a los mismos a pesar de tener una condena firme (o medida cautelar) de prohibición de aproximación o entrada en ellos y/o de acercarse a cualquiera de sus trabajadores.

Y es que bien es sabido que en muchas ocasiones las medidas cautelares y/o las sentencias firmes de condena de aquellos delincuentes reincidentes que suelen cometer delitos contra la propiedad en establecimientos MERCADONA, por ejemplo, incluyen la prohibición de aproximación o entrada en los mismos durante un cierto periodo de tiempo. No obstante, el problema que ello planteaba para la compañía era que, salvo que tales individuos fueran reconocidos por el personal de la tienda, seguían accediendo sin control alguno a los supermercados de la cadena y delinquiendo una y otra vez, con el elevado coste que ello representaba para la empresa y con el riesgo que, en muchas ocasiones, suponía para sus trabajadores.

Ante tal problemática, no obstante, en julio de 2020 la mencionada compañía decidió anunciar[590] que iba a empezar a instalar en sus establecimientos cámaras de seguridad dotadas de un *software* de reconocimiento facial capaz, como se ha dicho, de detectar, entre los rostros de todas las personas que accedieran a los mismos, aquellas caras cuyos datos biométricos coincidieran con los de las imágenes introducidas en su base de datos, pertenecientes a todas las personas sobre las que pesaran medidas cautelares vigentes de prohibición de aproximación/entrada al establecimiento o aproximación a sus trabajadores o condenas firmes en el mismo sentido, estableciendo de inmediato un contacto directo con la central receptora de alarmas para que el personal de seguridad pudiera así llamar a la policía y advertir al responsable de la tienda de la presencia del infractor.

Desde el Departamento de Seguridad de MERCADONA afirmaban, no obstante, que el sistema era altamente garantista y respetuoso con los derechos fundamentales y únicamente registraba los datos biométricos de aquellas personas que resultaban identificadas, siendo el resto borrados en décimas de segundo, y que se introducía en la base de datos la información biométrica de la persona sometida a medida cautelar o condenada en firme solamente en caso de que en el Auto o sentencia se incluyera la autorización judicial a MERCADONA para que, a través de los medios tecnológicos existentes, hiciera cumplir la prohibición impuesta. Así, aunque se captaran los datos biométricos de todas las personas que accedieran a los supermercados que contaran con dicho sistema, solamente se tratarían aquellos que identificaran a individuos sometidos a orden de alejamiento o

[590] Véase Fernandez, 2020.

condena firme de prohibición de entrada y/o aproximación de las tiendas MERCADONA o alguno de sus trabajadores.

Una de las ventajas del uso de tales sistemas de IA era que la introducción de los datos biométricos de la persona sobre la que recaía la medida cautelar o condena firme no requería de la realización de una fotografía de su rostro tras la decisión judicial, ya que el personal experto de MERCADONA, que tiene funcionando en todos sus establecimientos cámaras de seguridad de altísitma calidad que captan las imágenes de todo aquél que entra y sale (y las mantiene durante 30 días), realizaba un pantallazo de la imagen del individuo en cuestión, lo adjunta al expediente judicial y, posteriormente, de allí extraía los datos biométricos y los introducía en la base de datos del sistema de *software* de reconocimiento facial para que pudieran detectarse futuros incumplimientos.

En virtud de lo expuesto, mediante la utilización de la herramienta de IA empleada por MERCADONA, la instrucción de los delitos de quebrantamiento de medida cautelar y condena por incumplimiento de la prohibición de entrada en sus establecimientos o aproximación a sus trabajadores quedaba resuelta en tiempo récord (y con máximo ahorro de recursos personales y materiales), aumentaban exponencialmente las posibilidades de disuadir de la comisión de nuevos delitos contra la propiedad en tales supermercados, y se incrementaba, sin duda, la seguridad de los empleados, habida cuenta de que el sistema conectaba con el personal de seguridad de forma automática, en cuanto detectaba la presencia de un infractor.

En virtud de ello, cada vez han sido más los Juzgados y Tribunales que han impuesto a los individuos reincidentes de cometer delitos contra la propiedad en los establecimientos MERCADONA medidas cautelares de prohibición de aproximación o entrada en ellos o a sus trabajores o condenas en el mismo sentido, lo que sin duda ha contribuido a hacer posibles todas las utilidades que el sistema ofrece. Así, entre otras, la sentencia 20/18, de 16 de enero, dictada por el Juzgado de lo Penal nº1 de Granada y la Sentencia 159/2018, de 1 de abril, dictada por el Juzgado de lo Penal nº1 de Plasencia, impusieron a individuos reincidentes una condena que incluía la prohibición de entrada en la totalidad de los supermercados MERCADONA de España (aproximadamente unos mil seiscientos), con autorización a la compañía para, a través de los medios tecnológicos existentes, hacer cumplir lo dispuesto en la sentencia.

Para poner tales sistemas en funcionamiento, la mencionada compañía empezó a hacer pruebas en uno de sus establecimientos cerrados al público ubicado en Valencia. Allí, durante un tiempo, se testaron distintas cámaras y distintos sistemas de *software* de los proveedores más punteros del mercado y, para ello, diversos trabajadores, que decidieron ceder voluntariamente sus datos biométricos, intentaron ponerlos a prueba de diversas formas: accediendo con gafas, mascarilla, gorra, mano en parte de la cara, etc, lo que hizo que la empresa acabara decantándose por el sistema ofrecido por la antedicha compañía tecnológica israelí AnyVision, ya que fue el que mejores resultados arrojó.

Tras ello, tal y como me expuso Daniel Larios Caparrós, la compañía decidió instalar las primeras cámaras dotadas de tales sistemas de IA en sus cuarenta establecimientos ubicados en Mallorca, habida cuenta de que era el territorio de España donde contaban con más prohibiciones judiciales de aproximación o entrada en MERCADONA (alrededor de setenta), así como en una tienda en Valencia (base de la compañía) y en cuatro tiendas en Zaragoza (donde también tenían varias prohibiciones judiciales de aproximación o entrada a su favor).

A pesar de lo expuesto, sin embargo, la Agencia Española de Protección de Datos (AEPD) inició una investigación de oficio que culminó el 2 de julio de 2021 con un informe desfavorable[591], al entender que la propuesta de tratamiento de datos basados en el reconocimiento facial con fines de identificación presentada por la cadena de supermercados no estaba autorizada de acuerdo con lo dispuesto en el artículo 9.2.g) del RGPD, carecía de base de legitimación al amparo de lo previsto en el artículo 6.1 de dicho cuerpo legal y era contraria a los principios de necesidad, proporcionalidad y minimización, siendo, por ende, un tratamiento ilícito, lo cual conllevó no solo un revés jurídico sino admeás una elevada multa de dos millones y medio de euros para Mercadona que fue abonada de forma voluntaria el 19 de julio de 2021.

Ya con anterioridad, no obstante, la justicia, en concreto la Audiencia Provincial de Barcelona (Sección 9º), ya se había pronunciado de modo desfavorable al respecto por medio de Auto 72/2021 (Rec. 840/2021), de 15 de febrero.

[591] Véase AEPD, 2021.

La mencionada resolución -en la que se expresó claramente que se trataba de un tema complejo que suscitaba muchas dudas a nivel jurídico- vino a confirmar el Auto dictado el 27 de Septiembre de 2019 por el Juzgado de lo Penal nº 24 de Barcelona que denegó la autorización a Mercadona para utilizar medios automatizados de captación de datos biométricos de los penados en orden a detectar su entrada en cualquier establecimiento de dicha cadena radicado en Cataluña. En relación con ello, en el antedicho Auto de la Audiencia Provincial se concluyó que el uso de sistemas de reconocimiento facial para los mencionados fines no tenía cabida en lo dispuesto en el RGPD, que con carácter general prohíbe en su artículo 9 el tratamiento de los datos biométricos, salvo en contadas excepciones (entre otras, que sea "*necesario por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros, que debe ser proporcional al objetivo perseguido, respetar en lo esencial el derecho a la protección de datos y establecer medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del interesado*"[592]). Y en tal sentido se puso de manifiesto, por un lado, que en el caso de Mercadona no se protegía el interés público, sino los intereses privados de la compañía; por otro lado, que la empresa estaba vulnerando los derechos y libertades del resto de personas que accedían al supermercado sin prohibición de entrada alguna, puesto sus datos biométricos eran tratados de forma ilegítima (aunque se detectaran los rostros buscados en 0,3 segundos y los demás se borraran automáticamente); y, finalmente, que la medida no resultaba proporcional ni idónea y que se carecía de ley específica que amparara el reconocimiento facial.

Tales conclusiones jurídicas, no obstante, en mi opinión (que con toda probabilidad no coincide con la de la AEPD, que es muy restrictiva en todo caso), resultarían muy distintas si el sistema de reconocimiento facial empleado de forma privada por Mercadona fuera utilizado de forma pública por las autoridades competentes con fines de prevención, detección, investigación y enjuiciamiento de infracciones penales o de ejecución de sanciones penales (por ejemplo, mediante la instalación de cámaras en los puntos de acceso de una población para detectar la presencia de personas con prohibiciones de entrada por la comisión de ciertos ilícitos penales). Y es que, tal y como se verá más adelante, el tratamiento de datos biométricos dirigidos a identificar de manera unívoca a una persona

[592] Artículo 9.g.) del RGPD.

física con los fines de prevención, investigación, detección de infracciones penales, incluidas la protección y la prevención frente a las amenazas contra la seguridad pública, queda expresamente permitido en el artículo 13.2 de la LO 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales. No obstante, como es lógico, desde luego en cualquier caso resultaría necesario estudiar el escenario concreto para poder determinar la legitimidad de la medida.

Y, finalmente, procede hacer referencia a aquellas ocasiones en las que el juez de instrucción impone a las personas investigadas en una causa judicial la obligación realizar comparecencias *apud acta* en el Juzgado que tramita la misma o en el más cercano a su domicilio uno o varios días al mes, con el fin de intentar asegurar su localización para la práctica de diligencias de investigación pendientes o la para la citación al acto del plenario, entre otras.

Como consecuencia de ello, no obstante, los juzgados de guardia de toda España reciben a diario a multitud de personas que acuden a firmar y cumplir así con la obligación judicialmente impuesta, lo cual resulta altamente ineficiente. Y es que, por un lado, ello obliga a los funcionarios de Justicia a atender a todo aquél que acude a hacer la mencionada gestión, con el tiempo y el empleo de recursos que ello conlleva (habida cuenta de que en muchas ocasiones, además, debe remitirse el documento de firma a los juzgados que tramitan las causas); y, por otro lado, tal sistema tradicional dificulta muchísimo el control del cumplimiento de la medida, habida cuenta de que es el propio funcionario encargado del caso el que normalmente tiene que ir comprobando que se lleven a cabo las comparecencias en los días indicados para, en caso contrario, dar cuenta al juez por si entiende oportuno agravar la medida, lo cual en la práctica suele ser difícil y caótico, habida cuenta de la gran carga de trabajo que hay en los juzgados.

No obstante, si fueran instaladas en los juzgados de instrucción de toda España cámaras dotadas con sistemas de IA reconocimiento facial que incluyeran en sus bases de datos imágenes de todas aquellas personas con obligación de comparecer *apud acta* ciertos días al mes, tales ineficiencias podrían solventarse. Y es que ello, por un lado, permitiría a los ciudadanos simplemente acudir a un punto de control ubicado en la puerta de cualquier Juzgado de instrucción del país (tipo el que existe en los aeropuertos para el control de

aduanas), que detectaría sus datos biométricos para, en cuestión de segundos, dar por cumplida su obligación de comparecencia *apud acta*; y, por otro lado, existiría la posibilidad de que tal información, que debería ser volcada en una base de datos nacional, hiciera saltar la alarma en casos de incumplimiento, lo que permitiría a los funcionarios encargados de los casos detectar a los incumplidores simplemente accediendo a diario al sistema (o incluso, en caso de que este fuera más sofisticado, bastaría con acceder al expediente digital para verificar el historial de comparecencias *apud acta* realizadas por cada individuo -lugar, fecha, hora etc-, que podría alertar de forma automática en casos de incumplimiento).

De tal modo, pues, se ahorrarían sin duda recursos materiales y personales (previa inversión que, no cabe duda a la larga, resultaría rentable) en la gestión y control de las obligaciones de comparecencia *apud acta*.

Además, resulta interesante apuntar que, en los casos expuestos, una opción a contemplar sería la de conectar el sistema de IA de reconocimiento facial de los puntos de control de los juzgados de instrucción con una base de datos en que se hallaran incluidas las imágenes de todas aquellas personas que tuvieran vigente una requisitoria judicial, de forma que enseguida que un inidviduo se personara para cumplir con su obligación de comparecencia *apud acta*, en caso de hallarse además en busca y captura, saltara una alarma que conectara directamente con el personal de seguridad del juzgado y/o con la policía, de modo que este pudiera ser informado de sus cuentas pendientes y, en su caso, detenido. Y es que ahora, son los propios funcionarios del juzgado los que, si detectan que la persona que acude a firmar tiene vigente una orden de busca y captura, tienen que ingeniárselas para (en la mayoría de casos con falta de personal y espacio) llamar a los vigilantes de seguridad y a la policía para informar de la situación sin que el individuo se percate, con el fin de intentar garantizar el éxito de la operación, lo cual resulta lento, desagradable y complejo y en muchas ocasiones, peligroso e infructuoso, habida cuenta de que este acaba poniéndose nervioso y marchando antes de que llegue la fuerza pública.

a.2.7) Identificación y detección de posibles delincuentes que cruzan la frontera de nuestro país

Para muchos de nosotros no es ajena la existencia en los aeropuertos de máquinas de control automático de entrada o salida del país a través del pasaporte, a las que ya se ha hecho referencia en puntos anteriores. Así, tales dispositivos, generalmente de autoservicio, requieren para su uso insertar tal documento en un lector y mirar de frente a una cámara que, si detecta que somos la misma persona que la que consta como titular, abre la barrera para dejarnos pasar sin más dilación (lo que generalmente ahorra largas filas de viajeros).

En relación con ello es interesante mencionar que desde el 29 de junio de 2009, fecha en que entró en vigor el Reglamento (CE) nº444/2009 del Parlamento Europeo y del Consejo, de 28 de mayo de 2009, por el que se modificó el Reglamento (CE) nº2252/2004 del Consejo, sobre normas para las medidas de seguridad y datos biométricos en los pasaportes y documentos de viaje expedidos por los Estados miembros, todos los pasaportes emitidos por Estados Miembros de la UE son biométricos y, por ende, contienen nuestros datos de tal clase (tanto faciales como de huellas dactilares). Y, asimismo, es importante advertir que en virtud de lo dispuesto en el Reglamento (UE) 2016/399 del Parlamento Europeo y del Consejo, de 9 de marzo de 2016, por el que se establece un Código de normas de la Unión para el cruce de personas por las fronteras (el denominado Código de fronteras Schengen), los Estados miembros están obligados a realizar controles sistemáticos de todos los viajeros que entran y salen del país con el fin de intensificar la seguridad principalmente debido a la seria amenaza terrorista que acecha a Europa.

De acuerdo con ello por una parte, por ejemplo, en algunos aeropuertos franceses (entre otros, Paris-Charles de Gaulle y Paris-Orly, Lyon, Marsella y Niza) y estaciones de tren (entre otras, Estación del Norte de París)[593] se emplea el sistema PARAFE, una función automatizada de autoservicio que emplea tecnología biométrica para ayudar a los viajeros (siempre que sean ciudadanos del espacio Schengen) a cruzar la frontera de forma más rápida e independiente. Tal servicio, de uso voluntario, ofrece dos opciones: la de autenticación facial (pasaporte, tarjeta de embarque y rostro), que tarda entre 10 y 15 segundos en identificar al viajero, y la de autenticación de huellas digitales (pasaporte, tarjeta de embarque y huella digital), que tarda 30 segundos.[594]

[593] Gobierno de Francia, 2020.
[594] Aeropuerto de París, s.f..

En España, por otra parte, tal y como ya se ha expuesto en puntos previos, algunas fronteras (entre otras, las de los aeropuertos de Madrid, Barcelona, Gerona, Menorca, Mallorca, Alicante, Málaga y Tenerife, las del Puerto de Algeciras y el paso terrestre entre La Línea de la Concepción y Gibraltar) cuentan también con los mencionados sistemas de autenticación automática (conocidos como "*Automated Border Control*" -ABC-) a través del pasaporte o DNI electrónico[595], que se hallan conectados con bases de datos que detectan personas que tienen requisitorias policiales y/o judiciales pendientes, prohibiciones de entrada o salida en el país, etc, así como el uso de documentos robados o los intentos de usurpación de identidad. Y es que tales casos, el sistema hace saltar una alarma, tal y como ya se ha expuesto con anterioridad, y conecta de forma inmediata con el puesto fronterizo de Policía Nacional, quien se encarga de gestionar el asunto *in situ*, lo cual resulta muy útil a efectos judiciales.

Sin embargo, tal y como también se ha expuesto anteriormente, la conexión con dichas bases de datos únicamente se realiza con la información personal contenida en el pasaporte presentado, sin hacer uso de los datos biométricos del viajero, que solo se emplean para comprobar si este es la misma persona que la titular de tal documento oficial de identificación. Ello, no obstante, considero que resulta insuficiente, a efectos de investigación penal, puesto que si el volcado de información que las máquinas automáticas realizan en las bases de datos contuviera ya *ab initio* los datos biométricos del viajero, en cuestión de segundos este podría resultar identificado, sin dejar margen alguno a posibles fugas y ahorrando trabajo (no siempre fácil) a los agentes fronterizos, que tienen que lidiar con innumerables asuntos en su día a día con recursos limitados.

Y es que imaginemos, por ejemplo, que un ciudadano español investigado en un procedimiento por tráfico de drogas sobre el que pesa una medida cautelar de prohibición de salida del territorio nacional decide quebrantar tal medida y emplear el pasaporte de un familiar parecido a él para huir del país a un territorio de dentro del espacio Schengen, por ejemplo, Alemania. Imaginemos, asimismo, que tal individuo se dirige al aeropuerto y, en vez de simplemente mostrar el documento al personal de la línea aérea (al que seguramente se le pasarían por alto las pequeñas diferencias físicas entre el titular y el portador del

[595] Véase Ministerio del Interior, 2008.

mismo), tiene que introducirlo en una máquina de control automático que, en cuestión de segundos, detecta a través de sus datos biométricos que, por un lado, no es la misma persona que la que dice ser (con la consiguiente imputación de un delito de usurpación de identidad) y, por otro lado, lo identifica de modo inmediato y advierte que además tiene vigente una prohibición de salida del territorio nacional (con la consiguiente imputación de un delito de quebrantamiento de medida cautelar), por lo que hace saltar una alarma que conecta de forma inmediata con el puesto fronterizo y hace que los agentes de policía procedan a su rápida detención.

Así, vemos que mediante el uso de un sistema de reconocimiento facial (y de huellas dactilares) a través de IA, un caso que podría haber incluso pasado desapercibido, se convierte en una causa judicial de fácil instrucción, que seguramente, terminaría con una sentencia de conformidad, con el consiguiente ahorro de recursos personales y materiales que ello conlleva para la Administración de Justicia.

a.2.8) Identificación de cadáveres a través de sistemas de IA de reconocimiento facial que emplean técnicas de superposición craneofacial

Es importante poner de manifiesto que la tecnología de IA de reconocimiento facial puede ser empleada por los médicos forenses para identificar cadáveres mediante técnicas de superposición craneofacial.

En relación con ello, es interesante hacer referencia al sistema Skeleton-ID, que surgió a raíz de una colaboración llevada a cabo entre la compañía de I+D en IA "Panacea Cooperative Research" y la Universidad de Granada[596] y se ha convertido en el primer *software* del mercado que "*automatiza gran parte del proceso de identificación forense mediante técnicas de antropología física*"[597], empleando fundamentalmente dos técnicas de IA: el aprendizaje automático o *Machine Learning* y la visión por computador o *Computer Vision*.[598]

[596] Tras más de catorce años de investigación, dos proyectos europeos, cinco proyectos nacionales y siete tesis doctorales
[597] Álvarez, 2018.
[598] Véase más en Skeleton ID, s.f..

Y es que, tal y como afirma Óscar Ibáñez, uno de los creadores del mencionado sistema, actualmente la identificación forense se realiza de una manera "*eminentemente manual, lenta, subjetiva, y propensa a errores*".[599] No obstante, a través del mencionado *software*, dotado de un sistema de IA de reconocimiento facial que aportará "*rapidez, fiabilidad y objetividad*", podrá llevarse a cabo la denominada superposición craneofacial, es decir, el análisis de un cráneo hallado y la detección de las correspondencias anatómicas con la cara de aquella persona a la que se sospecha que puede pertenecer. Así, el mencionado investigador afirma: "*La potencialidad es muy grande porque lo único que necesita son varias fotografías lo más cercanas posibles al momento de la muerte*".[600]

No obstante, ya se está trabajando en la creación de un nuevo sistema todavía más fiable, basado en la comparación de radiografías, y es que como indica el antedicho investigador "*la ventaja radica en que comparas directamente la misma cosa, un hueso con un hueso*" (…) "*En la superposición cráneofacial, comparamos el cráneo con la cara, y hay un grado de incertidumbre que es el tejido blando*".[601]

Y es que, si bien actualmente las técnicas de identificación de cadáveres más utilizadas son los análisis genéticos y dactiloscópicos, en casos de grandes catástrofes (atentados terroristas, accidentes aéreos, desastres naturales, etc) estas pueden resultar ineficientes e incluso, inviables, habida cuenta de su elevado coste económico, por lo que el mencionado *software* de reconocimiento facial podría resultar de gran utilidad. No obstante, tal y como se afirma por sus creadores, el médico forense siempre deberá estar detrás del caso y tendrá la última palabra, al menos, por el momento.

a.2.9) Localización de autores y víctimas de delitos cometidos a través de la red de Internet, tales como trata de seres humanos, pornografía infantil, corrupción y abuso de menores e incapaces etc.

[599] Diario de León, 2018.
[600] *Idem.*
[601] *Idem.*

Es comúnmente conocido que en múltiples ocasiones, delitos tales como la trata de seres humanos, la pornografía infantil o la corrupción y el abuso de menores se cometen o dejan rastro en la red de Internet, lo cual puede resultar de gran utilidad para los investigadores.

Así, no es infrecuente encontrar anuncios *on line* que tratan de engañar y reclutar a mujeres, por ejemplo, para acudir a un determinado país con un fin legítimo (entre otros, con una oferta de trabajo), a pesar de que estas luego, al llegar, son obligadas a prostituirse; o hallar imágenes de menores de edad en actitudes sexuales explícitas compartidas por los usuarios de Internet a través de las redes sociales u otras vías de transmisión de información *on line*; o advertir la presencia de imágenes de menores o incapaces siendo víctimas de abusos sexuales por parte de adultos que, asimismo, circulan por Internet.

En virtud de ello, equipos de expertos de los distintos cuerpos de policía del mundo entero se dedican de forma exclusiva a rastrear la red y a buscar exhaustivamente información que les permita obtener datos sobre los presuntos autores de tales delitos y sus víctimas, no solo con el ánimo de proteger a estas últimas sino también con el fin de lograr el castigo de los delincuentes, ya que los hechos suelen ser de extrema gravedad. Ello, no obstante, no siempre resulta tarea fácil, habida cuenta de la inmensa cantidad de imágenes y anuncios que circulan a diario por la red, a velocidad extrema. Y es que, como sabemos, la capacidad humana para analizar información es limitada, mucho más que la de una "máquina" entrenada para ello.

Además, es interesante poner de manifiesto que cuerpos policiales de Canadá y EEUU están empleando un sistema de reconocimiento facial diseñado por la controvertida compañía Clearview AI (que maneja una base de datos de tres mil millones de imágenes) para identificar a menores víctimas de delitos de abuso sexual.[602] Y es que, según exponen los investigadores, el uso de tal *software* está posibilitando conocer la identidad y la localización de los niños y adolescentes que aparecen en fotografías o vídeos de explotación sexual colgados en la red, lo cual no solo resulta fundamental para identificar posteriormente a los autores de tales delitos y proceder a su detención, sino también para

[602] Kashmir & Dance, 2020.

filiar a las víctimas y ofrecerles personarse en los procedimientos, siendo que su testimonio puede resultar fundamental para el éxito de las investigaciones.

Como consecuencia del laborioso y eficaz trabajo que desarrollan los *software* descritos, el Abogado General de Manhattan llegó a poner de manifiesto que en un año las investigaciones de tráfico de personas habían ascendido de treinta a trescientos, lo que no es de extrañar, puesto que, por ejemplo, el sistema XIX es capaz de elaborar un informe completo que aporte información sobre una posible red de tráfico sexual en tan solo seis horas.[603] Así, la utilidad de este tipo de sistemas en investigaciones tan arduas y complejas, por la cantidad de fuentes y datos que se manejan, resulta evidente e indiscutible, no solo para el éxito de las mismas sino también para el ahorro de recursos personales policiales y judiciales, que pueden ser redistribuidos y destinados a otros asuntos que requieran la atención humana.

Y es que, imaginemos que llegan a manos de la policía unas imágenes de pornografía infantil que circulan por la red y muestran, de forma explícita, la imagen de tres menores manteniendo sexo con un adulto. Asimismo, imaginemos que la policía inicia una investigación para identificar tanto al autor del delito de distribución de pornografía infantil, como al autor del delito de abuso sexual de menores y a las víctimas de ambos, que en este caso coinciden. Para ello, no obstante, imaginemos que el grupo de expertos en este tipo de delitos se dedica a hacer, entre millones de fuentes y datos, una búsqueda exhaustiva de imágenes en la red en las que aparezcan tales sujetos y de conexiones de usuarios que las comparten. Pensemos que, al cabo de los días, logran descubrir que un usuario ha compartido dichas imágenes en un portal de Internet, por lo que piden autorización judicial para identificar al titular de la dirección IP desde donde se ha llevado a cabo la distribución, pero siguen sin tener pista alguna sobre la identidad de las personas que aparecen en las fotografías. Imaginemos que, finalmente, logran dar con la persona que compartió las imágenes en la red, la detienen y la ponen a disposición judicial, pero esta, o bien se niega a declarar, o bien asegura al juez que no tiene conocimiento alguno de la identidad de aquellos individuos que se ven en las imágenes (lo cual, seguramente, sea verdad). En tal caso, lo más probable es que acabe juzgándose el delito de pornografía infantil pero no el delito de abuso sexual, que quedaría impune, lo cual es extremadamente

[603] Kashmir & Dance, 2020.

grave, puesto que posiblemente los niños que aparecen en las imágenes sean víctimas recurrentes de alguna mafia que se dedica a cometer delitos contra la libertad sexual de los menores.

No obstante, mediante el uso por parte de la policía de un *software* de reconocimiento facial potente y confiable, en cuestión de horas se podrían rastrear millones de páginas web, plataformas *on line*, redes sociales etc con el fin de buscar imágenes que contuvieran el rostro del autor de los hechos o de las víctimas, lo que probablemente, en un corto periodo de tiempo, podría proporcionar a la policía información valiosísima, desde el origen de tales imágenes hasta la identificación del autor y las víctimas sometiendo sus rostros a examen con los introducidos en una vasta base de datos.

a.2.10) Fines de prevención y frustración de planes delictivos

La India, a través de su colaboración con la empresa israelí Cortica, dispone uno de los sistemas de cámaras de seguridad con tecnología de reconocimiento facial más punteros del mundo.

Y es que las miles de cámaras instaladas en la vía pública de tal país no solo cuentan con sistemas de reconocimiento facial que permiten identificar a los viandantes y a los vehículos en tiempo real, sino que además tienen la capacidad de captar y analizar micro expresiones faciales y detectar patrones de comportamiento, haciendo saltar alertas que se transmiten a las autoridades en casos de prever la comisión de un delito, siendo aparentemente capaces de detectar, incluso, cuándo, por ejemplo, en un mercado va a desencadenarse alguna acción violenta organizada o cuándo, en un templo, va a producirse un ataque terrorista. Además, las mencionadas cámaras de seguridad están conectadas con bases de datos militares y gubernamentales y, asimismo, tienen como finalidad la búsqueda activa de personas huídas o fichadas por la policía que puedan resultar una amenaza para la paz del país.[604]

[604] Véase Barbieri, 2019.

De acuerdo con lo expuesto, tal tecnología, que está ya proliferando en otros países[605], permitiría sin duda que las fuerzas policiales se adelantaran y evitaran la comisión de ciertos delitos que podrían causar un daño irreparable, permitiendo además generar indicios que podrían ser empleados posteriormente en la instrucción de las causas seguidas, principalmente, por las tentativas frustradas (o, incluso, por delitos perfeccionados tales como, por ejemplo, pertenencia a organización criminal, entre otros).

b) Reconocimiento de voz

b.1) *Concepto*

A modo de concepto puede decirse que el reconocimiento de voz es aquella tecnología que permite, a través de la IA, la identificación de personas y/o la comprobación de su identidad a través de sus datos biométricos, en este caso, los sonidos producidos por la vibración de sus cuerdas vocales[606] (la voz).

Los sistemas de reconocimiento de voz, no obstante, pueden multiplicar de forma exponencial su utilidad (especialmente en el ámbito de la investigación criminal) en caso de resultar combinados con sistemas de IA que empleen técnicas de Procesamiento del Lenguaje Natural (PLN), a las que se hará referencia más adelante con más profundidad, capaces de recibir y comprender mensajes y ejecutar comandos hablados, si bien es importante dejar claro que la función de identificación de un individuo por su voz es distinta a la de la comprensión de lo que este está diciendo.

Así, procede hacer referencia a los sistemas de identificación o autenticación de voz.

Y es que, al igual que el resto de datos biométricos a los que se ha venido haciendo referencia a lo largo de la presente Sección, la voz es una característica fisiológica (depende fundamentalemente de la forma y el tamaño de la boca y la garganta, la nariz, la longitud de la laringe, el peso corporal, etc y da lugar al timbre y tono natural de cada uno) y

[605] Corea del Sur, por ejemplo, pretende tener instaladas en 2022 tres mil cámaras dotadas de tal tecnología en distintos puntos estratégicos de la ciudad de Seúl.
[606] RAE.

conductual (depende del acento, el tipo de lenguaje empleado, la variedad del vocabulario, etc) de la persona que resulta singular y única, por lo que se antoja idónea para poder llevar a cabo tareas de identificación y verificación de identidad.

No obstante, procede poner de manifiesto que, mientras que las huellas dactilares, el ADN y los rasgos faciales tienen la consideración de datos biométricos estáticos, tal y como ya se ha expuesto con anterioridad, la voz es reputada un dato biométrico dinámico, puesto que va sufriendo variaciones considerables con el paso del tiempo a causa de diversos factores (envejecimiento, estrés, drogas, alcohol, enfermedades, etc), por lo que su potencial identificador presenta grandes desafíos.

Hasta ahora, el análisis y reconocimiento de la voz humana se ha venido llevando a cabo en el ámbito de la investigación criminal, generalmente, por expertos forenses (o incluso por las propias víctimas o los testigos) mediante técnicas tradicionales, si bien en no pocas ocasiones, sobre todo en casos de voces muy similares (por ejemplo, de miembros de la misma familia) o de grabaciones de baja calidad, las limitaciones han quedado más que patentes, puesto que resulta muy difícil para el oído humano detectar según qué pequeñas diferencias o patrones que, no obstante, pueden resultar fundamentales para resolver un caso con éxito.

Así, por ejemplo, en 1986 David Shawn Pope, un ciudadano estadounidense, fue condenado a pena de prisión por la supuesta comisión de un delito de agresión sexual. Dicha condena se basaba, entre otras pruebas, en el reconocimiento de su voz realizada por la víctima en el acto del juicio y en la confirmación efectuada por un experto perito que aseguró haber detectado su "huella vocal" en unos mensajes que había dejado, presuntamente, en el contestador automático de la mencionada mujer. No obstante, quince años después, el antedicho ciudadano fue exculpado y liberado de prisión, puesto que una prueba de ADN descartó su participación en los hechos, lo que, sin duda, es una muestra de las nefastas e irreparables consecuencias que puede acarrear la incorrecta identificación de un presunto delincuente, en este caso a través de su voz.[607]

[607] Véase Michigan State University, 2012.

Ello, no obstante, puede quedar solucionado (o, al menos, minimizado) con las nuevas herramientas de IA de reconocimiento de voz que se están desarrollando desde comienzos del siglo XXI, ya que mediante técnicas de aprendizaje automático o *Machine Learning* estas están siendo entrenadas para conseguir unos grados de precisión inalcanzables para el cerebro humano y dotar de mayor objetividad al proceso mediante la eliminación (o, al menos, disminución) de la subjetividad inherente al ser humano.

Y es que, los mencionados *software* de reconocimiento de voz contienen algoritmos ejercitados a base de conjuntos masivos de datos vocales que les permiten extraer patrones y determinar la forma de hablar que tenemos los humanos, lo que luego posibilita la observación y detección de aquellas características o desviaciones que puede tener una voz concreta que hacen que sea distinta del resto, lo cual resulta fundamental para la determinación de la identidad de la persona que hay detrás.

No obstante, tal y como afirma Carlos Delgado Romero, Jefe del Laboratorio de Acústica Forense de la Comisaría General de Policía Científica del Cuerpo Nacional de Policía: "*un sistema de reconocimiento automático no es un ente autónomo capaz de alcanzar, por sí mismo, un criterio de identificación. Sencillamente nos encontramos ante una herramienta inteligente de parametrización y modelización de los sonidos del habla, con una poderosa capacidad de clasificación*." Y es que, prosigue: "*Hoy en día, el grado de interacción científico-máquina resulta determinante en diferentes etapas del proceso, principalmente a la hora de seleccionar los audios objeto de comparación y al interpretar las puntuaciones proporcionadas por el sistema*.", añadiendo que "*Existen muchos niveles de información a los que estos sisteas todavía no tienen acceso o este es muy limitado*. *A título de ejemplo, cabría citar la detección de una impostación, la posible influencia de de factores exógenos o endógenos en el habla (alcohol, drogas, patologías, etc*); *la evaluación de la información lingüística que tiene que ver con las funciones del lenguaje; la estimación de la competencia, usos o hábitos lingüísticos del locutor; la apreciación y comparación de los aspectos sociolectales, geolectales, idiolectales, etc. Este tipo de valoraciones corresponden al especialista quien, también, en clara primacía respecto a la*

máquina, es capaz de adjudicar y conjugar factores de tipicidad, consistencia o adecuación.". [608]

Ya en 2015, un estudio [609] encargado al Doctor Geoffrey Stewart Morrison, de la Universidad de Aston (Birmingham, Reino Unido), por la Oficina para Asuntos Legales de Interpol, reveló que el uso de sistemas de reconocimiento automático de voz estaba afianzado en el entorno policial.

Más recientemente, en el Simposio 21 de la Red Europea de Institutos de Ciencias Forenses (ENFSI) y del Working Group for Forensic Speech and Audio Analysis (FSAASWG), celebrado los días 12 y 13 de septiembre de 2019 en Budapest (Hungría), se presentó un informe que puso de manifiesto que, de veinte institutos de la ENFSI, catorce empleaban sistemas de reconocimiento automático de voz (en concreto, nueve empleaban el sistema BATVOX; tres empleaban el sistema *Nuance Forensic*; dos empleaban el sistema *Ivocalise*; uno empleaba *Asis*; y otro empleaba *Phonexia*).[610]

Dicha tecnología, si bien está más extendida para usos comerciales (por ejemplo, Google introdujo en 2015 el sistema de autenticación de voz denominado Trusted Voice en Google Play[611] [612]), como era de esperar, ya está siendo empleada en el campo de investigación criminal.

Así, por un lado, por ejemplo, los servicios de inteligencia y la policía británica al parecer emplearon dicho sistema de IA para intentar dar con la identidad del supuesto miembro del Estado Islámico que en 2014 decapitó al periodista James Foley, habida cuenta de que su voz era el único dato contenido en el vídeo que se publicó que podía aportar alguna información sobre su identidad. Así, según lo asegurado por Elizabeth McClelland, analista forense de voz, al portal de noticias de la CNN, las autoridades británicas trataron de cotejar

[608] Delgado, 2020, págs. 62-63.
[609] Stewart & otros, 2016, págs. 92-100.
[610] ENFSI & FSAASWG, 2019, pág. 69.
[611] El usuario graba su voz en el teléfono móvil diciendo "OK Google" y los patrones de esta quedan guardados para identificar al interesado cada vez que quiera acceder al dispositivo.
[612] Véase Padla, 2015.

la voz del presunto yihadista con las miles de grabaciones contenidas en sus bases de datos de audio.[613]

Además, según se desprende de diversos documentos clasificados (de 2004 a 2012) hechos públicos[614], la Agencia de Seguridad Nacional norteamericana (NSA) y el FBI ya llevan más de una década empleando este tipo de sistemas para identificar a personas sospechosas de haber cometido actos delictivos. Y es que, a partir de tales documentos se puede comprobar que los analistas de la NSA, en materia militar, por ejemplo, en el marco de la operación IRAQI FREEDOM, confirmaron a través de un sistema de reconocimiento del habla que la voz que parecía de Saddam Hussein era realmente la suya, al contrario de lo que algunos creían inicialmente; y, en materia de terrorismo, dichos analistas crearon la huella de voz del conocido terrorista Osama Bin Laden a través de las transmisiones y los vídeos en los que este aparecía, así como la de Abu Musab al-Zarqawi, otro alto mando de Al-Qaeda.[615]

Por otro lado, entidades financieras como JP Morgan Chase & Co ya emplean sistemas de IA de reconocimiento de voz con el fin de prevenir y detectar fraudes y estafas. Así, tal y como se desprende de su página web[616], dicho banco hace uso de la aplicación de IA "Voice ID", una función de seguridad que permite crear una "huella de voz" de cada uno de sus clientes, única e intransferible, a partir del análisis de más de cien características distintas (tales como la forma de la boca, el tracto vocal, el tono y el acento), e identificarlos así, de forma rápida, cuando llaman a su centro de servicio al cliente, facilitando un acceso seguro a su cuenta. Con tal fin, mediante una llamada telefónica inicial (siempre y cuando el cliente preste su consentimiento) se crea su mencionada "huella de voz", de modo que la próxima vez que llame, simplemente con su habla se verifica rápidamente su identidad y se reduce la necesidad de responder a preguntas de seguridad adicionales[617], lo que evita y detecta usurpaciones de identidad y posibles fraudes. Desde la antedicha entidad bancaria se asegura que, con carácter general, Voice ID funciona incluso si el cliente tiene la voz

[613] Véase Smith-Park, 2014.
[614] Véase Kofman, 2018.
[615] Véase más Snowden Archive - The SID Today Files, 2006.
[616] JPMorgan Chase Bank, N.A., s.f.
[617] No obstante, en caso de que las condiciones de escucha no sean óptimas, por existir demasiado ruido de fondo, por ejemplo, puede optarse por hacer preguntas de seguridad adicionales.

alterada por padecer alguna enfermedad que implique cambios leves (en casos más severos, no obstante, se opta por verificar la identidad mediante preguntas de seguridad).

Y es que tal tecnología está proliferando cada vez más en el ámbito financiero, siendo varias las compañías que en los últimos tiempos la han introduciendo como servicio de seguridad para sus clientes (entre otras, cabe destacar, por ejemplo, el sistema "Voice Verification", empleado por la entidad Wells Fargo, con los mismos fines[618]).

Y, finalmente, en el marco del denominado proyecto SIIP ("*Speaker Identification Integrated Project*")[619], desarrollado entre 2014 y 2018 por un consorcio internacional de diecinueve socios (que incluían usuarios finales, tales como Interpol, Metropolitan Police -Reino Unido- o Carabinieri -Italia-; industrias, tales como Airbus o Nuance; Pymes, tales como Saillabs Technology; y representantes del mundo académico, tales como las Universidades de Warwik y Groningen), coordinado por la compañía Verint Systems Inc.[620], se creó el sistema que lleva el mismo nombre y que permite tanto identificar rápidamente las voces de sospechosos (ayuda además a identificar el género, la edad, el idioma y el acento y, asimismo, capta las clonaciones de voz), como detectar y aislar aquellas conversaciones de interés que puedan surgir en un contexto de intervenciones telefónicas. Y ello es de gran utilidad, ya que durante el desarrollo del mencionado proyecto, por ejemplo, se consiguió con éxito la identificación de terroristas mediante un motor de identificación del habla que funcionaba en diferentes fuentes, tales como Internet, a través de las redes sociales, y teléfonos fijos y móviles.

La mencionada herramienta de IA, sin duda prometedora, funciona del siguiente modo: cada vez que un agente de policía introduce una referencia de voz en el sistema, este se encarga de eliminar los ruidos ambientales y de potenciar la voz, que es entonces empleada para crear una referencia única. Posteriormente, el sistema busca en las bases de datos policiales locales y en la base de datos global de Interpol posibles coincidencias con referencias de voz de criminales ya fichados (cuya voz consta almacenada en el sistema)

[618] Véase Wells Fargo, s.f.
[619] Véase Interpol, s.f.
[620] Una empresa de inteligencia procesable con sede en Nueva York.

y, asimismo, busca en las redes sociales vídeos, grabaciones, etc para hallar potenciales coincidencias con voces de personas que no constan fichadas por la policía.

Una de las grandes ventajas del mencionado sistema SIIP es, sin duda, la posibilidad que brinda a cuerpos de policía de todo el mundo[621] para acceder a la la base de datos de voz global de Interpol, tanto para hacer consultas como para introducir nuevas voces, lo que supone la creación de un fichero que cada cada vez será más extenso y, por tanto, útil, y que sin duda hará que la calidad de tal herramienta de IA continúe mejorando y evolucionando. Y es que la idea es que, en el futuro, la identificación del habla se combine con otras tecnologías de reconocimiento biométrico (tales como el reconocimiento de huellas dactilares y el reconocimiento facial), y se logre así facilitar y agilizar las identificaciones y verificaciones de identidad de los delincuentes, nacionales e internacionales, a través de todas las bases de datos interconectadas.

En España, actualmente, los distintos cuerpos policiales ya emplean sistemas automáticos de reconocimiento de voz. El más extendido, por su alta calidad y vanguardia, es sin duda el sistema BATVOX[622], de origen nacional (empleado, entre otros, por el Cuerpo Nacional de Policía). No obstante, procede poner de manifiesto que, a pesar de ser una de las herramientas más punteras del mercado, todavía necesita un elevado grado de interacción con los expertos en varios momentos del proceso de reconocimiento de voz.

Así, por un lado, los mencionados expertos juegan un rol fundamental en el filtrado de las muestras de voz indubitadas que se introducen en la bases de datos del sistema, puesto que este requiere unos estándares de calidad muy altos para poder arrojar resultados precisos con éxito, por lo que hace falta una tarea previa de selección de aquellas muestras con mayor potencial identificativo, en función de su formato y características. Posteriormente, la labor de la herramienta es esencialmente automática, ya que esta procede a detectar patrones en la muestra de voz introducida y crear un modelo concreto del individuo al que pertenece, a partir de las características únicas que se desprenden de ella. A continuación, una vez nutrida la base de datos, los expertos proceden a solicitar a la "máquina" que compare las muestras de voz dubitadas (cuantas más se tengan, mejores resultados se

[621] Aquellas pertenecientes a los ciento noventa y cuatro países que actualmente son miembros de Interpol.
[622] Véase Scientific Analytical Tools, s.f.

obtendrán) de que disponen con aquellas indubitadas que se contienen en ella, con el fin de detectar coincidencias. Respecto de esto último procede decir, no obstante, que las comparaciones deben hacerse siempre entre muestras que tengan las mismas características, es decir, que correspondan a un individuo del mismo género y que las grabaciones de voz tengan las mismas propiedades. Así, finalmente, el sistema arroja una lista de posibilidades (candidatos), en función de los *match* conseguidos, que será analizada por el equipo de expertos.

En la actualidad, no obstante, y a pesar de la gran precisión del mencionado sistema, este suele emplearse por los cuerpos policiales únicamente como complemento al trabajo humano, para verificar o descartar resultados obtenidos por los expertos en ciertos casos, ya que los elevadísimos estándares de calidad exigidos, que requieren que las muestras de voz tengan una serie de características no siempre fáciles de conseguir, hacen que sea útil únicamente en alrededor del 25% de los casos, tal y como me manifestó el Doctor Carlos Delgado Romero, Jefe del Laboratorio de Acústica Forense de la Comisaría General de Policía Científica del Cuerpo Nacional de Policía. Y es que los expertos cuentan actualmente con otros métodos alternativos, asimismo eficaces, que les permiten trabajar con muestras de menor calidad, lo cual les resulta mucho más sencillo. No obstante, los sistemas de reconocimiento de voz como el mencionado han evolucionado mucho en los últimos años y, sin duda, van a seguir haciéndolo, siempre con tendencia a la autonomía y a la automatización, por lo que habrá que ir siguiéndolos muy de cerca para detectar el gran abanico de posibilidades que van aportando, aunque en palabras del mencionado Doctor, siendo que la voz es un dato biométrico dinámico (a diferencia de las huellas dactilares y el ADN), el éxito de los sistemas de IA siempre va a ser limitado, ya que van a necesitar un experto humano que ayude detectar e interpretar los posibles cambios (voluntarios o involuntarios) que puedan haber.

Dicho lo expuesto, únicamente sistemas con estándares de calidad de las muestras más bajos, que a su vez permitieran la parametrizacion, la creación de modelos de voz indubitados y su posterior análisis y comparación con muestras dubitadas, supondrían un avance significativo en la ciencia forense del reconocimiento de voz. En relación con ello, no obstante, hay que decir que no es buena idea abusar de aquellas herramientas capaces de eliminar elementos disturbadores de la voz, como por ejemplo, ruidos ambientales,

puesto que a la vez que se lleva a cabo tal operación, se pierde calidad y potencial identificativo del habla, ya que es imposible disociar una cosa de la otra, por lo que tales tareas son difíciles de compatibilizar con el reconocimiento de voz.

b.2) ***Posibles utilidades en la instrucción de las causas***

Además de lo ya expuesto, procede poner de manifiesto que los sistemas de reconcimiento de voz a través de IA pueden tener diversas utilidades en el ámbito de la investigación criminal.

b.2.a) Aumento de la eficiencia y la eficacia en el análisis de las intervenciones telefónicas mediante la identificación de los intervinientes y, en su caso, el reconocimiento de palabras clave

Bien es sabido que una de las tareas más arduas y exigentes de la policía es la de la escucha y el análisis de las conversaciones telefónicas intervenidas, habida cuenta de la gran cantidad de recursos humanos que deben emplearse para el éxito de la diligencia. Y es que en la mayoría de ocasiones, los agentes de policía hacen turnos para no perderse ni un segundo de las conversaciones telefónicas mantenidas por los presuntos delincuentes, con el fin de de obtener la máxima información posible que resulte de utilidad en la investigación.

Tal tediosa tarea, no obstante, podría ser aliviada con el uso de sistemas de IA de reconocimiento de voz entrenados para, por un lado, reconocer las voces de los individuos que intervinieran en las conversaciones y, por otro lado, incluso (mediante técnicas de procesamiento de lenguaje natural (PLN) que, tal y como ya se ha advertido con anterioridad, serán analizadas más adelante con más detalle), reconocer palabras clave y, en caso de ser pronunciadas, hacer saltar la alarma para dar aviso a los agentes.

Así, resultaría muy útil un sistema de reconocimiento de voz que distinguiera y disgregara por sí solo las distintas voces (dubitadas) que fueran interviniendo en las conversaciones y, de forma automática, las "lanzara" contra las bases de datos existentes para tratar así de hallar coincidencias con las voces indubitadas contenidas en ellas. Y es que, de tal forma,

no solo se conseguiría detectar qué palabras debieran ser imputadas a un interviniente u otro, sino también averiguar (o, en su caso, confirmar o descartar) sus identidades.

Además, el hecho de que el sistema estuviera entrenado para reconocer ciertas palabras clave y, en caso de ser escuchadas, ejecutar la orden de hacer saltar la alarma, permitiría que los agentes no tuvieran que permanecer ininterrumpidamente en la sala de escuchas, ya que únicamente tendrían que acudir en el momento en que sonara dicha alerta (o en los casos menos urgentes, siendo que ya que el minuto exacto de la generación de la alerta quedaría registrado, podrían simplemente escuchar la grabación a partir de tal momento, con el consiguiente ahorro de tiempo).

No obstante, si bien lo anterior devendría altamente útil para ahorrar recursos policiales y aumentar la eficiencia de la investigación, lo cierto es que en muchos casos no resultaría de gran ayuda, al menos por el momento. Y es que, por un lado, procede advertir que actualmente las bases de datos de voz no cuentan con un registro tan amplio de voces como para arrojar resultados de mucho éxito; y, por otro lado, es bien sabido que con frecuencia los delincuentes hablan en "jerga" propia de su actividad o, si sospechan que pueden estar siendo escuchados, emplean palabras o expresiones clave que solo un profesional humano, a base de oír horas y horas de conversaciones, puede llegar a comprender.

b.2.b) Identificación de los verdaderos autores de los hechos

En ocasiones, la prueba estrella de la comisión de ciertos delitos es una grabación. Esta suele provenir de conversaciones mantenidas en directo, por teléfono o por Internet, y al ser presentada al juez de instrucción puede dar lugar a diversos escenarios.

En primer lugar, procede poner de manifiesto que si bien, por un lado, es posible que la persona interviniente en la conversación, a la que se le imputa la comisión de un delito, reconozca que esa es su voz y simplemente discuta que los hechos sean constitutivos de infracción penal, por otro lado, es posible que esta niegue haber participado en ella. Y es entonces cuando llega el problema, ya que aun en los casos más obvios hay que echar mano a los informes periciales de expertos, en ocasiones de elevado coste, para que aporten sus conclusiones al respecto. No obstante, mediante el uso de un sistema de reconocimiento de voz a través de la IA, tales casos podrían quedar resueltos de forma mucho rápida y precisa.

Así, imaginemos que una víctima de un delito de estafa aporta grabaciones en las que se puede escuchar cómo la autora de los hechos la engaña para tratar de obtener una transferencia bancaria de un elevado importe y obtener así, de forma injusta, un beneficio patrimonial. Pensemos, asimismo, que la mencionada víctima conoce a la perfección a dicha estafadora, siendo que es una vieja "amiga", si bien esta, al ser llamada a declarar por la policía y posteriormente por el juez de instrucción, niega ser la persona que interviene en las conversaciones aportadas. Ante ello, figurémonos que se acuerda una prueba pericial judicial para determinar si efectivamente la voz de las grabaciones corresponde a la persona investigada y, con tal fin, se emplea un sistema de IA de reconocimiento de voz que compara ambas voces y arroja un resultado más que concluyente de forma inmediata, confirmando que se trata de la misma persona.

En tal caso, de no haber empleado dicho sistema de IA, lo más seguro es que, entre la designación del perito judicial, la aceptación del cargo y la elaboración del informe hubieran transcurrido como mínimo tres meses, a los que habría que sumar mínimo dos más hasta que se diera traslado a las partes y al Ministerio Fiscal, se aportara la contraprueba pericial de la defensa y esta se proveyera, por lo que el uso de tal tecnología sin duda habría ahorrado varios meses de instrucción.

Además, con bastante probabilidad, en caso de haberse empleado un sistema de IA potente y fiable, la presunta autora de los hechos hubiera procedido a reconocer su participación de los mismos, pudiendo así reducir al mínimo la instrucción de la causa y, en su caso, transformar el procedimiento en Diligencias Urgentes y terminarlo con una sentencia de conformidad.

Y lo mismo serviría en el caso contrario, puesto que de no ser la persona investigada la misma que intervino en las conversaciones aportadas, el sistema sería capaz de detectarlo de forma inmediata, lo que permitiría sobreseer sin dilación la causa frente a esta.

En segundo lugar, procede poner de manifiesto que en ocasiones el/los autor/es de los hechos es/son un/os perfecto/s desconocido/s para la víctima, que únicamente ha tenido acceso a su voz.

Así, imaginemos un caso de delito de amenazas telefónicas en que un adolescente recibe llamadas desde un número oculto, de forma reiterada, escuchando una voz al otro lado del teléfono que le profiere expresiones tales como "*si te veo por la calle te mato*", "*si no dejas quinientos euros debajo de la maceta de tu portería mañana, te doy una paliza*", o "*más te vale hacerme caso porque si no te rajaré con una navaja*". Pensemos que tal chico explica a sus padres lo que está sucediendo y estos deciden interponer una denuncia en la policía, si bien no pueden aportar dato alguno sobre la identidad del autor porque la desconocen. Figurémonos, no obstante, que el menor tiene grabadas las conversaciones y las aporta como prueba, de modo que la policía decide introducir la voz del delincuente en la base de datos del sistema de IA de reconocimiento de voz para averiguar si esta coincide con alguna de las que se hallan allí introducidas, arrojando un resultado positivo, siendo que la voz del autor de los hechos ya constaba registrada por haber sido condenado por varios delitos de amenazas en el pasado.

En tal caso, de no haber empleado el mencionado sistema de IA con casi total seguridad el caso hubiera quedado sobreseído por desconocimiento de autor, con la sensación de inseguridad y angustia que ello hubiera ocasionado al menor y a sus familiares, y de esta forma, sin embargo, podría quedar resuelto de forma rápida y eficaz (siempre, por supuesto, otorgando a la defensa la posibilidad de contradicción en el plenario).

Resulta evidente que, al menos por el momento, las posibilidades de éxito del uso de este tipo de sistemas de IA en casos como el expuesto son escasas (de hecho, prácticamente inexistentes), habida cuenta de que no se cuenta con bases de datos de voz suficientes y capaces de aportar valor de forma significativa a la investigación de las causas, pero cierto es que si se les empieza a nutrir de información de calidad, en un futuro no muy lejano podrían arrojar resultados muy útiles y esperanzadores.

No obstante, en relación con ello es importante hacer referencia a IdentiVox, un *software* "*desarrollado íntegramente en España capaz de comparar la grabación de una voz con miles de otras en cuestión de minutos*"[623], que ya ha venido siendo empleado para resolver casos reales en múltiples ocasiones desde hace más de una década.

[623] Abad, 2015.

Así, tal y como afirma Carlos Delgado, jefe del Laboratorio de Acústica Forense del Cuerpo Nacional de Policía, dicho *software* fue empleado para investigar y averiguar la identidad de los autores del atentado terrorista perpetrado en la Terminal 4 del Aeropuerto de Madrid el 30 de diciembre de 2006: "*El atentado de la T4 reflejó la madurez de los sistemas automáticos, que nosotros veníamos usando desde el año 97. Hasta entonces se trabajaba normalmente solo con lingüistas, pero nuestro laboratorio apostó por ese software*".[624] No obstante, el mencionado agente pone de manifiesto la problemática a la que se enfrentaron él y su equipo: "*Aquella llamada dubitada [así llaman los forenses las muestras de voz de cuyo autor desconocen] era óptima en calidad, pero el problema era obtener la grabación del sospechoso para compararla*. (…) "*Una patrulla rural de la Guardia Civil había detenido a dos individuos en Guipúzcoa en los primeros días de enero. La cercanía en el tiempo con el atentado hizo sospechar que estaban relacionados con él*", si bien finalmente destaca: "*La sentencia lo resaltó: la prueba de voz fue relevante no solo para identificar a uno de los autores, sino también para calificar el delito como terrorista*".[625]

b.2.c) Detección de voces artificiales o sintéticas (especial referencia al caso de los "DeepFakes")

El progreso tecnológico, en el ámbito del habla, ha traído consigo una revolución sin precedentes: la creación del habla sintética o artificial, es decir, generada por una máquina (generalmente, un ordenador), que ciertamente cada vez se parece más al habla humana, habiendo llegado a niveles tan precisos que en muchos casos dificulta la distinción entre una y otra.

En tal sentido, tal y como afirma el ya mencionado Doctor Carlos Delgado Romero: "*El último peldaño superado, en la escalera de síntesis del habla, es el de la clonación de la voz. A partir de tan solo unos minutos de la grabación de la voz de un sujeto, un algoritmo de reconocimiento es capaz de modelar su cualidad vocal (la voz de Pedro, de María o de Juan). Una vez generado el modelo del locuto, el sistema está en disposición de hacer uso de él para reproducir cualquier mensaje escrito o hablado que le sea transferido*".[626] Y es

[624] Abad, 2015.
[625] *Idem.*
[626] Delgado, 2020, pág. 73.

que ello, entre otros, ha dado lugar al denominado fenómeno de los "*DeepFakes*", imágenes (generalmente de vídeo) manipuladas, con contenido falso.

Y es que tal aplicación del habla artificial, sin duda, genera múltiples oportunidades para los delincuentes, que pueden emplearla para cometer delitos de injurias o calumnias, amenazas y estafas, por ejemplo, o para difundir mensajes terroristas, entre otros, mediante la usurpación de la identidad de otro, siendo que ello facilita la ocultación de su verdadera identidad.

Ello, lamentablemente, supone un nuevo reto para la policía científica, que se las tiene que ingeniar para, principalmente, aprender a detectar tales casos. En tal sentido, el Doctor Delgado asegura que "*El laboratorio de policía científica de España ya ha tenido que afrontar sus primeros casos con grabaciones de voz artificial. Los primeros requerimientos han estado relacionados con la determinación del carácter artificial o natural de los registros hablados*."[627], lo cual sin duda es un nuevo terreno a explorar por los expertos forenses, que ya se han puesto manos a la obra para buscar y encontrar "*las claves en el estudio de los patrones de articulación y coarticulación, de los rasgos suprasegmentales, de las funciones del lenguaje o, simplemente, del contexto sociolectal del discurso: plano expresivo, variaciones diafásicas o diastráticas*."[628]

No obstante, tal ardua tarea sin duda puede verse mitigada mediante el uso de la IA, habida cuenta de que ya se cuenta con potentes sistemas destinados a detectar casos de *DeepFakes* de forma rápida y precisa. Así, por ejemplo, Microsoft, con el fin de combatir el creciente problema de la desinformación, presentó en su blog oficial el 1 de septiembre de 2020 una nueva herramienta denominada Video Authenticator que permite detectar audios (e imágenes[629]) que hayan sido manipulados a través de la IA (*DeepFakes*).[630]

c) Reconocimiento de emociones

[627] Delgado, 2020, pág. 74.
[628] *Idem.*
[629] Por lo que también podría reputarse una utilidad aplicable a las herramientas de reconocimiento facial.
[630] Véase Burt & Horvitz, 2020.

c.1) *Concepto*

A modo de concepto puede decirse que el reconocimiento de emociones es aquella tecnología que permite, a través de la IA, la parametrización de expresiones, gestos, elementos del habla, etc de personas con el fin de detectar ciertos estados de ánimo, sentimientos o intenciones, mediante el análisis y, en su caso, la comparación, de sus datos biométricos, por lo general, del rostro y/o de la voz.

Dichos sistemas, habitualmente, con tales fines, cuentan con grandes bases de datos que albergan ingentes cantidades de información facial o de voz que servirán para, posteriormente, ser cotejadas por un algoritmo, previamente entrenado, con aquellas otras que puedan resultar de interés.

Si bien el análisis o el reconocimiento de emociones no es algo nuevo en el ámbito de la investigación criminal, puesto que todos estamos familiarizados, por ejemplo, con el conocido detector de mentiras o polígrafo, lo cierto es que hoy en día cobra especial relevancia por los enormes avances que, con el uso de la IA, está experimentando tal técnica.

Así, hoy en día, múltiples empresas, desde los gigantes tecnológicos como Google o NEC hasta compañías más pequeñas, como Eyeris, Affectiva o WeSee, están inviertiendo en tal tecnología.

Respecto de esta última firma británica, por ejemplo, es interesante advertir que asegura que su tecnología de IA puede detectar comportamientos sospechosos simplemente leyendo señales faciales que son imperceptibles para un ojo inexperto.

Su Director Ejecutivo, David Fulton, manifestó a la cadena de noticias BBC que WeSee era capaz de determinar el estado mental o la intención de un individuo a través de sus expresiones faciales, postura, gestos y movimientos utilizando únicamente secuencias de video de baja calidad, por lo que "*En el futuro, las cámaras de video en la plataforma de una estación de metro podrían usar nuestra tecnología para detectar comportamientos sospechosos y alertar a las autoridades sobre una posible amenaza terrorista. Y lo mismo*

se podría hacer con las multitudes en eventos como partidos de fútbol o mítines políticos."[631]

No obstante, el uso de tal tecnología pone en alerta a las organizaciones *pro* derechos humanos y a los expertos en protección de datos y, en tal contexto, Oliver Philippou, un experto en videoviligancia de la compañía IHS Markit, expuso sus temores al mencionado portal de noticias: "*Cuando se trata simplemente de identificar rostros, hay márgenes de error decentes: las mejores empresas afirman que pueden identificar a las personas con una precisión del 90% al 92%. No obstante, cuando también se intentan evaluar las emociones, el margen de error aumenta significativamente*."[632]

En el ámbito de la UE varias iniciativas del programa de investigación e innovación Horizon2020 exploran el uso de la misma, a saber: por un lado, el proyecto "*The Automatic Sentiment Analysis in the Wild*" (SEWA), que utiliza tecnología de reconocimiento facial de emociones para mejorar la comprensión automatizada del comportamiento humano interactivo[633]; y, por otro lado, el proyecto "*Intelligent Portable Border Control System*" (iBorderCtrl) ha diseñado un sistema de seguridad fronteriza automatizada, que incluye tal tecnología.

En relación con este último, tal y como se pone de manifiesto en su página web, procede poner de manifiesto que es un proyecto que tiene como objetivo permitir un control fronterizo más rápido y exhaustivo para aquellos nacionales de terceros países que crucen las fronteras terrestres de los Estados miembros de la UE y que, para ello emplea tecnologías que van desde la verificación biométrica, hasta la detección automática de engaños y la evaluación de riesgos.

Con el fin de llevar a cabo la mencionada detección automática de engaños se emplea el denominado "Automatic Deception Detection System" (ADDS) que realiza, controla y evalúa la entrevista previa al registro que un Avatar efectúa a cada viajero, cuantificando la probabilidad de engaño medinte el análisis de los microgestos no verbales de los

[631] Thomas, 2018.
[632] *Idem.*
[633] Véase Automatic Sentiment Analysis in the Wild (SEWA), s.f..

entrevistados (sin perjuicio de que cada caso sea verificado adicionalmente por parte de un agente humano). [634]

En España, por su parte, la Policía Nacional ya cuenta con el sistema de IA denominado "Layered Voice Analysis" (LVA), adquirido a la empresa israelita Nemesysco, cuya tecnología patentada está basada en un conjunto de parámetros vocales que se relacionan con las emociones humanas clave, con el fin de poder detectar las señales emocionales ocultas en un discurso e identificar las potenciales intenciones de engaño.

Tal y como se afirma en la página web de la compañía[635], tales parámetros vocales se identifican a partir de un banco de archivos de audio tomados en diferentes idiomas y en numerosos entornos, incluidos interrogatorios policiales y centros de llamadas. Así, el sistema emplea un algoritmo para detectar diferentes tipos de patrones y anomalías en el flujo del habla y clasificarlos en términos de estrés, excitación, confusión y otros estados emocionales relevantes.

Asimismo, en nuestro país, la Secretaría de Estado de Seguridad, dependiente del Ministerio del Interior, ha encargado a la empresa Herta Security un sistema de IA, denominado BioObserver, para uso exclusivo de Guardia Civil, Policía Nacional y Centro Nacional de Inteligencia, que emplea tecnología de reconocimiento facial para analizar imágenes de video y detectar expresiones y gestos humanos relacionados con las emociones para proceder, así, a sacar conclusiones al respecto.

Tal y como se afirma en la página web de la compañía[636], BioObserver es un *software* no invasivo, ya que se basa en técnicas de procesamiento de imágenes, y es capaz de detectar emociones faciales básicas como la alegría, la tristeza, el enfado, el miedo, la aversión, la sorpresa y la neutralidad, así como microexpresiones más sutiles del rostro tales como el fruncimiento del ceño, el parpadeo, el levantamiento de cejas, el cambio de dirección de la mirada y la orientación de la cabeza, entre otros, lo cual puede resultar muy útil a los agentes policiales para avanzar en la investigación.

[634] Véase Community Research and Development Information Service (CORDIS), s.f..
[635] Nemesysco, s.f..
[636] Herta, s.f..

En relación con lo expuesto es importante poner de manifiesto que hoy en día todavía no existen sistemas de IA efectivos que combinen el reconocimiento de emociones simultaneando el análisis facial y de voz, habida cuenta de que ello requeriría de procesadores extremadamente potentes que, al menos por el momento, no son sencillos de crear, si bien en el futuro ello puede llegar a convertirse en una opción muy útil y completa.

No obstante, lo que sí que hacen actualmente los distintos cuerpos policiales es emplear, de forma complementaria, todos los sistemas que tienen a su disposición, tanto de reconocimiento de emociones a través del rostro como de reconocimiento de emociones a través de la voz, para sacar a así resultados más concluyentes.

En relación con ello, especialmente llamativo es el hecho de que, tal y como denunciaron las ONGs ProPublica y Wired en 2019, cientos de escuelas, hospitales y otros lugares públicos de todo el mundo (incluidos más de cien en EEUU) emplean un sistema de IA denominado "detector de agresión"[637] creado por una empresa llamada Louroe Electronics, que funciona con la colocación de micrófonos en áreas públicas y el uso de ordenadores para escuchar y emitir alertas cuando se detecten voces humanas que se consideren indicativas de agresión.

No obstante, ante el potencial de vulneración de derechos y libertades que esta nueva tecnología ostenta además de las ya mencionadas, la voz crítica de las organizaciones *pro* derechos humanos, especialmente de EEUU, no se han hecho esperar.

Y es que, desde la organización ACLU, se denuncia la invasión de la privacidad que el uso de tal sistema entraña, habida cuenta de que en muchas ocasiones es mucho más invasivo que un dispositivo capte la voz que la imagen, siendo que ello da acceso al contenido de las conversaciones. Además, por un lado, se alega que el sistema no incluye tecnología de reconocimiento de voz para identificar quién está hablando en una grabación, si bien este podría combinarse con sistemas de reconocimiento facial o con grandes bases de datos de huellas de voz para permitir que el sistema Louroe no solo sepa lo que se dice, sino también quién lo dice. Y, por otro lado, se aduce que no está claro que pueda interpretarse de manera precisa y confiable (y menos por un ordenador) qué voces humanas indican agresión,

[637] En inglés, "aggression detector".

siendo que las emociones se expresan de formas que varían según la cultura, el individuo y la situación.[638]

c.2) ***Posibles utilidades en la instrucción de las causas***

c.2.a) Guía para la correcta interpretación de las declaraciones efectuadas en sede de instrucción

Si bien en los Cuerpos de Policía existen agentes expertos en realizar interrogatorios e interpretar los gestos, los movimientos, los tonos de voz, etc de las personas que prestan declaración, lo cierto es que estos deben contar con un nivel de formación muy elevado y una experiencia muy extensa para poder extraer conclusiones claras y precisas. No obstante, ni los fiscales ni los jueces de instrucción cuentan con conocimientos técnicos al respecto, siendo que se basan en su mera experiencia e intuición, así como en los criterios jurisprudencialmente establecidos (básicamente basados en la lógica de la declaración -coherencia interna- y en el suplementario apoyo de datos objetivos de corroboración de carácter periférico -coherencia externa-) [639] para valorar la credibilidad de las declaraciones.

De acuerdo con ello, resultaría enormemente útil para los investigadores poder emplear un sistema de IA que detectara posibles anomalías en las declaraciones de las personas de interés y les proporcionara así una guía o un punto de partida en un sentido u otro.

Y es que imaginemos que un testigo, en el seno de una investigación por un delito de sustracción de menores, declara que la última vez que los vio fue el viernes anterior, sobre las 17h, en un coche negro. Pensemos que, a ojos de la persona que les toma declaración, este testigo no levanta sospecha alguna, si bien el sistema de IA de reconocimiento de emociones capta una microexpresión en el rostro del testigo, al decir la hora, que está relacionada con el nerviosismo por las consecuencias que ello puede producir, de modo que la policía procede a poner en duda tal información y a abrir una línea de investigación al respecto. O, por el contrario, imaginemos que el propio agente que toma declaración al mencionado testigo ya intuye que este puede estar mintiendo respecto de tal información y

[638] Véase Bogus, 2019.
[639] Véase entre otras la STS, Sala 2ª, nº79/2016, de 10 de febrero.

ello viene corroborado por "la máquina", por lo que con más determinación, decide investigar en profundidad qué hay detrás de tal aseveración.

Ello, además, como queda documentado, puesto que el propio *software* emite un informe que puede quedar plasmado incluso en formato "pdf", podría servir de base para armar un buen relato para solicitar la adopción de medidas o diligencias de investigación al juez instructor, como por ejemplo, el volcado del teléfono o el registro del domicilio de tal testigo, que pasaría a resultar investigado, entre otras.

c.2.b) Guía para la correcta interpretación de las personas investigadas en los casos de reconstrucción de hechos

La diligencia de reconstrucción de hechos, si bien no consta regulada en la Ley de Enjuiciamiento Criminal, sí ha sido jurisprudencialmente admitida y desarrollada, a pesar de que su práctica es bastante remota y excepcional, ya que solo tiene lugar cuando el juez de instrucción considera necesario tener contacto directo con la presunta escena del crimen.

Tal diligencia, que puede ser acordada de oficio, a instancia del Ministerio Fiscal o de cualquiera de las partes personadas, suele llevarse a cabo en presencia de las personas investigadas, con asistencia de sus Letrados, si bien su intervención es voluntaria, con el fin de garantizar el derecho a la defensa (quedando a salvo los casos en que se acuerde *ope judicis* de forma motivada).

Durante la práctica de la mencionada diligencia, que suele consistir en una reproducción de los hechos lo más fiel posible a la realidad, con el fin de verificar y valorar las declaraciones prestadas y el resto de indicios o elementos de prueba disponibles, lo habitual es que tanto el juez de instrucción como el fiscal y los Letrados (así como la Policía Judicial, en su caso), traten de no quitar ojo a las personas de interés, ya que cualquier gesto, reacción o movimiento puede ser de utilidad para la instrucción de la causa.

Hoy en día, no obstante, la interpretación de tales emociones resulta limitada, puesto que, por un lado, ninguna de las personas presentes en la diligencia puede estar observando de forma permanente a los mencionados individuos de interés y, por otro lado, depende únicamente de los conocimientos y la experiencia que puedan tener quienes intervienen en

el reconocimiento de hechos (lo cual, además, da cabida a la subjetividad), por lo que sin duda resultaría enormemente útil para los investigadores poder emplear un sistema de IA que detectara posibles anomalías en las actuaciones de los sujetos participantes y les proporcionara así una guía o un punto de partida en un sentido u otro.

Y es que imaginemos que un presunto homicida, en el seno de una reconstrucción de hechos por un delito de homicidio de un menor cuyo cuerpo, no obstante, no ha aparecido, al pasar por una determinada estancia de la vivienda donde presuntamente se cometió el crimen realiza una serie de gestos faciales, casi impercetibles para los humanos, que denotan tensión, nerviosismo y miedo. Tal circunstancia, desde luego, podría pasar desapercibida para los humanos que participan en la mencionada diligencia, puesto que, por un lado, como ya he dicho, es posible que estos no estuvieran manteniendo contacto visual con el investigado en ese preciso momento y, por otro lado, es probable que aunque lo hubieran tenido no hubieran captado el mensaje que esas microexpresiones faciales podían transmitir.

No obstante, en caso de emplear una cámara dotada con un *software* de reconocimiento de emociones que permitiera captar de forma constante todos los gestos, expresiones y movimientos de las personas de interés y, asimismo, grabarlas, sería sin duda de gran utilidad, puesto que no solo implicaría un examen más completo y preciso de tales circunstancias sino que además otorgaría la posibilidad (en especial a la defensa) de contradecir los resultados arrojados por el sistema, al constar toda la secuencia registrada en vídeo.

c.2.c) Incremento del valor probatorio de los informes de reconocimiento de emociones

Si bien en la actualidad los Cuerpos de Policía cuentan con agentes expertos en análisis de emociones, lo cierto es que pocas veces sus informes se presentan y son admitidos en el ámbito de la instrucción y del plenario, habida cuenta de la desconfianza que todavía existe frente a los mismos por distintos motivos (posible subjetividad, falta de precisión, etc).

No obstante, los sistemas de IA de reconocimiento de emociones, sin duda, pueden aportar un plus de valor a tales informes, puesto que les proporcionan un mayor grado de objetividad y de precisión.

Así, imaginemos un caso de denuncia falsa por violencia doméstica en que la acusación particular y el Ministerio Fiscal quieren aportar como prueba en el acto del plenario un informe pericial basado en el análisis de las emociones de la supuesta víctima (en este caso investigada) en su declaración en sede de instrucción. Pensemos, asimismo, que este únicamente está firmado por expertos de la Policía Judicial y es sometido a contradicción en el acto juicio, siendo finalmente desechado en la sentencia en virtud de las dudas que infunde la defensa.

Sin embargo, en caso de que tal informe hubiera sido ratificado no solo por los peritos autores del mismo sino que viniera avalado también por los resultados de un sistema de IA de reconocimiento de emociones (obviamente con certificado de calidad), este podría devenir una reforzada y objetivada prueba fundamental para la correcta resolución del caso.

d) Reconocimiento de huellas dactilares

d.1) *Concepto*

A modo de concepto puede decirse que el reconocimiento de huellas dactilares es aquella tecnología que permite, a través de la IA, la identificación de personas y/o la comprobación de su identidad mediante el análisis y, en su caso, la comparación, de sus datos biométricos, en concreto, las formas y proporciones de las crestas papilares de sus dedos, por lo general, de la mano.

Dichos sistemas, habitualmente, con tales fines, cuentan con grandes bases de datos que albergan ingentes cantidades de imágenes de huellas dactilares que servirán para, posteriormente, ser cotejadas por un algoritmo, previamente entrenado, con aquellas otras que puedan resultar de interés.

Generalmente, las huellas dactilares que tienen mayor utilidad a efectos de investigación ciminal son, por un lado, aquellas que se hallan en el lugar donde se han cometido los hechos delictivos, las denominadas “huellas latentes” o “marcas dactilares”, que se recogen a través de reactivos (principalmente químicos), habida cuenta de que, tal y como se pone

de manifiesto por Interpol[640], su cotejo o comparación con las huellas dactilares registradas en las bases de datos policiales puede llevar a identificar al/los autores de los hechos y a relacionar varios delitos entre sí; por otro lado, aquellas que se recogen a las personas detenidas, las denominadas "impresiones dactilares", que se captan a través del tintado de las huellas y su posterior plasmación en papel, o mediante un aparato con tecnología *Live Scan*, habida cuenta de que sirven para nutrir las bases de datos policiales y, asimismo, pueden luego ser cotejadas para identificar a presuntos criminales; y, finalmente, aquellas que pertenecen a personas que aparecen inidentificadas, con o sin vida, lejos del lugar del crimen pero que pueden relacionarse con el mismo, en calidad de víctimas o de autores, puesto que tienen potencial para guiar hacia la resolución de los casos investigados.

Principalmente, existen las siguientes técnicas (esencialmente de análisis y comparación de imágenes) para comparar algorítmicamente las huellas dactilares:

a) *técnicas basadas en la comparación de minucias o puntos singulares de la estructura de las crestas de las huellas dactilares*, que son las más extendidas a día de hoy (ya que son la versión automatizada del método empleado por las policías de todo el mundo en las investigaciones criminales) y consisten en el alineamiento de dos huellas dactilares con el fin de compararlas y hallar correspondencias entre el mayor número de minucias (puntos característicos de las crestas que incluyen las particularidades que se producen en su recorrido) posible[641];

b) *técnicas basadas en la estructura de las crestas*, que se basan en la comparación de la información completa de toda la estructura de las crestas de las huellas dactilares, teniendo en cuenta su grosor, los poros, la existencia de líneas albas, los cortes, etc. Así, a través de tales técnicas, se lleva a cabo la alineación de dos huellas dactilares con el fin de compararlas y hallar correspondencias entre la esctructura entera de sus crestas; y

c) *técnicas de correlación*, que se basan en la comparación de los píxeles de las imágenes de las huellas dactilares con el fin de detectar las eventuales correlaciones existentes entre

[640] Interpol, s.f.
[641] En España, los cuerpos policiales tienden a exigir la coincidencia entre al menos doce puntos para considerar que existe identidad entre dos huellas dactilares.

ellas bajo la premisa de que si existe un abrupto en la zona del nucleo de una cresta, por ejemplo, es muy probable que también exista en la de al lado.

Y es que, tal y como se pone de manifiesto por Interpol, las huellas dactilares son absolutamente genuinas, ya que no existen dos personas con tales datos biométricos iguales[642] y, además, estas no varían jamás (salvo en casos de destrucción por accidente o de modificación intencionada a través de cirugía plástica)[643], por lo que su análisis y comparación, especialmente a través de sistemas de IA que emplean algoritmos desarrollados para realizar tales tareas de forma automática, se erige como una de las técnicas más fiables y seguras en el ámbito de la identificación humana y, por ende, de la investigación criminal.

A pesar de lo expuesto, que hace referencia a las técnicas de reconocimiento de huellas dactilares más modernas que emplean IA, la dactiloscopia es una ciencia muy antigua. Así, ya en las antiguas Persia y Babilonia se utilizaban las huellas dactilares para autenticar registros en arcilla[644], y en el ámbito de la investigación delictiva, ya en China, sobre el año 300 d.C., se empleaban las huellas de las manos como prueba en los juicios por robo[645].

Desde entonces, y a lo largo de los siglos, diversos científicos han ido mostrando su interés y realizando hallazgos sobre el valor de las huellas dactilares. No obstante, no fue hasta el año 1880 cuando por primera vez se publicó[646] en una revista científica un artículo que puso de manifiesto la importancia de las huellas dactilares como elemento de identificación; y, posteriormente, en 1892 el científico británico Sir Francis Galton[647] fue el primero en publicar un libro completo sobre huellas dactilares, bajo el título de "*Finger Prints*"[648], en que demostró que estas son genuinas, distintas y únicas para cada ser humano. Ese mismo año, el antropólogo y policía argentino de origen croata Juan Vucetich, influenciado por los hallazgos del ya mencionado Sir Francis Galton, creó un método para verificar las coincidencias entre huellas dactilares y lo puso en práctica con éxito en un

[642] Ni tan siquiera los gemelos homocigóticos.
[643] Interpol, s.f..
[644] ECYT-AR, 2012.
[645] Gascueña, 2017.
[646] Por Sir William James Herschel, oficial del ejército británico en la India.
[647] Primo del conocido científico naturalista Charles Robert Darwin.
[648] Véase Galton, 1892.

caso de investigación de un homicidio de dos menores, habiendo descubierto, a través de una huella dactilar hallada en el lugar de los hechos, que la autora de este había sido la propia madre de los pequeños. Y, en 1907, la Academia de Ciencias de París reconoció de forma pública que el mencionado método de identificación humana creado por Juan Vucetich era el más efectivo y exacto que había existido hasta el momento.

En el ámbito español, las huellas dactilares han servido como fuente de filiación desde mediados del siglo XX, a raíz de la creación del denominado Documento Nacional de Identidad (DNI) en virtud de lo dispuesto en el Decreto de 2 de marzo de 1944 que exigía, para su expedición, entre otros datos personales, el dactilograma, en doble impresión, del índice derecho de la mano de cada persona.[649]

En la actualidad, los ciudadanos españoles debemos aportar nuestras huellas dactilares, a efectos meramente identificativos, para la expedición del DNI y del pasaporte. Y digo a efectos meramente identificativos porque hoy en día, en virtud de lo dispuesto en el RGPD y en la LOPD (que requieren que cualquier actuación relativa al tratamiento de datos personales quede circunscrita a su ámbito y finalidad específica) no se permite el acceso policial genérico a las bases de datos de huellas dactilares de dichos documentos oficiales para fines investigativos, puesto que para ello se hace únicamente uso de las bases de datos policiales de personas ya previamente "fichadas". No obstante, con autorización judicial, en casos concretos puede llegar a cotejarse una huella dactilar con utilidad criminal con la huella dactilar del DNI/pasaporte de un individuo determinado, si bien no existe la posibilidad de introducirla para cotejo en la base de datos de huellas dactilares de tales documentos puesto que esta no está dotada del sistema SAID (Sistema Automático de Identificación Dactilar), por lo que no es viable llevar a cabo una comparación automática (tendrían que ir buscándose coincidencias, una por una, con los millones de huellas dactilares que forman la base de datos nacional, lo cual es humanamente impracticable).

Respecto del DNI, hoy en día resulta de aplicación el Real Decreto 1553/2005, de 23 de diciembre, por el que se regula la expedición del Documento Nacional de Identidad y sus certificados de firma electrónica, modificado por el Real Decreto 414/2015, de 29 de mayo.

[649] Aunque con posterioridad las normas de expedición del DNI ampliaron las posibilidades del dedo a estampar para aquellos casos en que no resultara posible emplear el índice derecho.

En dicha legislación, por un lado, se establece la obligatoriedad de la expedición del DNI para todos los españoles "*mayores de catorce años residentes en España y para los de igual edad que, residiendo en el extranjero, se trasladen a España por tiempo no inferior a seis meses*"[650]; y, por otro lado, entre los requisitos para la expedición, se incluye la obtención de "*las impresiones dactilares de los dedos índices de ambas manos. Si no fuere posible obtener la impresión dactilar de alguno de los dedos o de ambos, se sustituirá, en relación con la mano que corresponda, por otro dedo según el siguiente orden de prelación: medio, anular o pulgar; consignándose, en el lugar del soporte destinado a tal fin, el dedo utilizado, o la imposibilidad de obtener alguno de ellos*."

Y, en la actualidad, en la UE contamos con los denominados DNIs electrónicos (DNIe), que permiten la identificación biométrica de sus titulares, a través del algoritmo "*Match on Card*", en puntos de acceso controlados.[651]

Y, respecto del pasaporte, resulta de aplicación el Real Decreto 896/2003, de 11 de julio, por el que se regula la expedición del pasaporte ordinario y se determinan sus características, modificado por el Real Decreto 411/2014, de 6 de junio, que en su artículo 10.5 establece: "*El pasaporte llevará incorporado un chip electrónico que contendrá la siguiente información referida a su titular: datos de filiación, imagen digitalizada de la fotografía, impresiones dactilares de los dedos índices de ambas manos, o los que en su defecto correspondan conforme al siguiente orden de prelación: medio, anular o pulgar.*"

Y es que en la actualidad, asimismo, en la UE contamos con los denominados pasaportes electrónicos, pasaportes biométricos o "ePassport", que contienen, tal y como se desprende del precepto transcrito, dos huellas dactilares y una fotografía digital, lo cual permite comprobar de forma muy fiable la titularidad de dicho documento a través de los datos biométricos de la persona que lo presenta como propio.

Así, tal y como ya se ha expuesto con anterioridad, en la mayoría de países de la UE se han incorporado infraestructuras biométricas en puertos y aeropuertos para el control de las

[650] Véase artículo 2 del Real Decreto 1553/2005, de 23 de diciembre, por el que se regula la expedición del Documento Nacional de Identidad y sus certificados de firma electrónica
[651] Véase más Cuerpo Nacional de Policía, 2020.

fronteras con el objetivo de que los escáneres de huellas dactilares y rasgos faciales identifiquen a los viajeros de una manera rápida y precisa, lo cual tiene una gran utilidad en materia de seguridad.

En España, actualmente, el sistema de reconocimiento automático de huellas dactilares a través de la IA empleado por nuestros cuerpos de policía es el Sistema Automático de Identificación Dactilar (SAID), de Interpol, que suele utilizar la mencionada técnica de las minucias (que en ocasiones combinan con otras, si bien el contenido de la mayoría de algoritmos permanece oculto, por lo que resulta imposible de saber) y, mediante el análisis y la comparación de imágenes lo que hace es aportar posibles candidatos (su número y el margen de error suele ser directamente proporcional a la calidad de la información) a los agentes de policía que han introducido la imagen de una huella dactilar dubitada, para que sean examinados por un experto en lofoscopia, que será quien haga el "*match*" final. Y es que dicho sistema permite registrar y cotejar las huellas dactilares recogidas en las reseñas efectuadas a los detenidos con las imágenes latentes recopiladas en el lugar de los hechos delictivos y, además, analiza las huellas latentes remitidas por otros cuerpos policiales a través de Sirene, Prüm, Europol e Interpol.

Y es que la mencionada base de datos internacional de huellas dactilares de Interpol denominada SAID resulta muy útil a los países miembros[652], ya que en caso de entender que puede existir alguna conexión internacional en los delitos que sus autoridades investigan, pueden proceder a cotejar, de forma rápida y efectiva, la información contenida en sus propias bases de datos nacionales de huellas dactilares con la del sistema SAID, que contiene más de doscientas veinte mil huellas registradas y más de diecisiete mil huellas latentes recogidas en lugares donde se han cometido delitos. [653]

Además, ya se está trabajando por dicha organización en la creación del denominado el Sistema Automático de Identificación Biométrica (SAIB), un nuevo sistema que permitirá acelerar las búsquedas y conseguir unos resultados más exactos, según se desprende de la propia página web de dicha organización policial.

[652] Actualmente, ciento noventa y cuatro.
[653] Véase Interpol, s.f..

En concreto, a nivel europeo existen diversos sistemas automáticos de identificación dactilar, que ya han sido mencionados con anterioridad, a saber: EURODAC, que debe analizarse a continuación con más detalle, tal y como se anunció en páginas previas, SIS II y VIS.

Respecto del sistema EURODAC ("*European Asylum Dactyloscopy Database*"), procede poner de manifiesto que se halla regulado por el Reglamento (UE) n°603/2013 del Parlamento Europeo y del Consejo, de 26 de junio de 2013, relativo a la creación del sistema «Eurodac» para la comparación de las impresiones dactilares para la aplicación efectiva del Reglamento (UE) n°604/2013, por el que se establecen los criterios y mecanismos de determinación del Estado miembro responsable del examen de una solicitud de protección internacional presentada en uno de los Estados miembros por un nacional de un tercer país o un apátrida, y de las solicitudes de comparación con los datos de Eurodac presentadas por los servicios de seguridad de los Estados miembros y Europol a efectos de aplicación de la ley[654]. Tal herramienta, gestionada por eu-LISA, se emplea para tomar decisiones de asilo por los veintisiete Estados Miembros de la UE más Reino Unido, Islandia, Liechtenstein, Noruega y Suiza, contiene una base de datos de huellas dactilares y emplea tal información para, por un lado, facilitar a dichos Estados la tarea de análisis de las solicitudes de asilo, mediante la comparación y comprobación de tales datos lofoscópicos,[655] y por otro lado, permitir que sus Cuerpos y Fuerzas de Seguridad efectúen consultas con fines de prevención, detección e investigación de delitos graves, entre ellos, terrorismo.

Y es que EURODAC cuenta con una base de datos biométricos que alberga las huellas dactilares de todos los solicitantes de asilo nacionales de fuera de la UE o del Espacio Económico Europeo (EEE) que, con los fines expuestos, pueden ser examinados y comparados de forma automática por parte de los Estados Miembros de la UE con las contenidas en una base de datos central.[656]

[654] Por el que se modifica el Reglamento (UE) n° 1077/2011, por el que se crea una Agencia europea para la gestión operativa de sistemas informáticos de gran magnitud en el espacio de libertad, seguridad y justicia.
[655] Véase más Comisión Europea, 2021.
[656] Véase más EUR-Lex, 2020.

Respecto del sistema SIS II, ya mencionado en páginas anteriores[657], en relación a la utilización de huellas dactilares, procede hacer especial alusión al Considerando 23 del Reglamento (UE) 2018/1862 del Parlamento Europeo y del Consejo, de 28 de noviembre de 2018, relativo al establecimiento, funcionamiento y utilización del Sistema de Información de Schengen (SIS) en el ámbito de la cooperación policial y de la cooperación judicial en materia penal, que dispone: "*La introducción de un servicio automatizado de identificación de impresiones dactilares en el SIS complementa el mecanismo existente de Prüm sobre acceso mutuo transfronterizo en línea a las bases de datos nacionales de ADN y a los sistemas automatizados de identificación de impresiones dactilares, tal como establecen las Decisiones 2008/615/JAI (1) y 2008/616/JAI (2) del Consejo. La consulta de datos dactiloscópicos en el SIS permite una búsqueda activa del autor de un delito. Por tanto, conviene prever la posibilidad de cargar los datos dactiloscópicos de un autor desconocido en el SIS, a condición de que la persona a quien correspondan esos datos dactiloscópicos pueda ser identificada con un muy alto grado de probabilidad como autor de un delito grave o un acto de terrorismo. Este es el caso, en particular, cuando se encuentran datos dactiloscópicos en un arma o en cualquier objeto utilizado para la comisión del delito. La mera presencia de los datos dactiloscópicos en el lugar del delito no debe considerarse como indicio de un muy alto grado de probabilidad de que dichos datos dactiloscópicos sean los del autor. Otro requisito previo para la creación de dicha descripción debe ser la imposibilidad de establecer la identidad del sospechoso a partir de datos procedentes de cualquier otra base de datos nacional, de la Unión o internacional pertinente. En caso de que una consulta de datos dactiloscópicos conduzca a una posible coincidencia, el Estado miembro debe proceder a verificaciones adicionales con la participación de expertos en la materia, a fin de determinar si el sospechoso es la persona a quien corresponden las impresiones dactilares almacenadas en el SIS, y debe establecer la identidad de la persona. El procedimiento debe estar sujeto al Derecho nacional. Tal identificación podría contribuir sustancialmente a la investigación y desembocar en una detención siempre que se cumplan todas las condiciones para proceder a la detención.*"

Por su parte, el Considerando 24 dispone: "*Los conjuntos completos o incompletos de impresiones dactilares o palmares encontradas en el lugar de un delito deben poder*

[657] Véase pág. 309.

cotejarse con los datos dactiloscópicos almacenados en el SIS, si puede acreditarse con un alto grado de probabilidad que pertenecen al autor de un delito grave o un delito de terrorismo, siempre que se haga también un cotejo de forma simultánea en las bases de datos nacionales pertinentes de impresiones dactilares. Se debe prestar particular atención al establecimiento de normas de calidad aplicables al almacenamiento de datos biométricos, en particular de datos dactiloscópicos latentes."

En el artículo 42 se establecen normas específicas aplicables a la introducción de fotografías imágenes faciales, perfiles de ADN y datos dactiloscópicos, y en relación a estos últimos se dispone, por un lado, que solo se introducirán en el sistema aquellos que cumplan las normas mínimas de calidad y las especificaciones técnicas, previa comprobación; y, por otro lado, que los datos dactiloscópicos introducidos en el SIS podrán consistir en una a diez impresiones dactilares planas y en una a diez impresiones dactilares rodadas, y en hasta dos impresiones palmares.

Y, finalmente, el artículo 43 dispone normas específicas aplicables a la comprobación o consulta mediante fotografías, imágenes faciales, perfiles de ADN y datos dactiloscópicos, y respecto a estos últimos establece que cuando una descripción del SIS disponga de ellos, estos "*se utilizarán para confirmar la identidad de una persona que haya sido localizada como consecuencia de una consulta alfanumérica realizada en el SIS.*", añadiendo que "*los datos dactiloscópicos del SIS en relación con descripciones introducidas con arreglo a los artículos 26, 32, 36 y 40 también podrán cotejarse utilizando conjuntos completos o incompletos de impresiones dactilares o palmares descubiertas en los lugares de comisión de delitos graves o delitos de terrorismo que estén siendo investigados, cuando pueda acreditarse con un alto grado de probabilidad que los conjuntos de impresiones pertenecen a un autor del delito y siempre que la consulta se realice de forma simultánea en las pertinentes bases de datos nacionales de los Estados miembros de impresiones dactilares*."

Y, respecto del sistema VIS, también mencionado ya en páginas anteriores[658], en relación a la utilización de huellas dactilares, procede hacer especial alusión a lo dispuesto en el artículo 15 del Reglamento (UE) nº767/2008 del Parlamento Europeo y del Consejo, de 9

[658] Véase pág. 312.

de julio de 2008 (Reglamento VIS), que hace referencia al examen de las solicitudes de visados y a la toma de decisiones relativas a las mismas, y establece que "*la autoridad competente en materia de visados correspondiente podrá efectuar búsquedas con uno o más datos de los siguientes* (…) *e) impresiones dactilares*.", así como a lo previsto en el artículo 20, que dispone que: "*Únicamente a efectos de identificar a cualquier persona que no cumpla o haya dejado de cumplir las condiciones de la entrada, estancia o residencia en el territorio de los Estados miembros, las autoridades responsables de los controles en los puntos de paso de las fronteras exteriores de conformidad con el Código de fronteras Schengen o de controlar en el territorio de los Estados miembros el cumplimiento de las condiciones de entrada, estancia o residencia en el territorio de los Estados miembros, tendrán acceso a la búsqueda con las impresiones dactilares de la persona.*" [659] y, finalmente, a lo determinado por el artículo 22 que, respecto del examen de las solicitudes de asilo y la toma de decisiones relativas a las mismas dispone que: "*Únicamente a efectos de examinar una solicitud de asilo, las autoridades competentes en materia de asilo tendrán acceso con arreglo al artículo 21 del Reglamento (CE) nº 343/2003 a la búsqueda con las impresiones dactilares del solicitante de asilo.*"

Asimismo, en el ámbito europeo procede hacer especial referencia a la ya citada Decisión 2008/615/JAI del Consejo, de 23 de junio de 2008, sobre la profundización de la cooperación transfronteriza, en particular en materia de lucha contra el terrorismo y la delincuencia transfronteriza (informalmente conocida como "Decisión Prüm"), en cuya virtud los Estados Miembros de la UE tienen la obligación de crear bases de datos nacionales de huellas dactilares (incluyendo índices de referencia procedentes de los sistemas automáticos de identificación dactilar -SAID- nacionales) que serán accesibles por los cuerpos policiales del resto de países para fines de investigación criminal (principalmente por delitos de terrorismo y delincuencia transfronteriza). Las consultas a las bases de datos deben llevarse a cabo para casos específicos con la finalidad de encontrar de forma automática coincidencias entre huellas dactilares que, en caso de darse, facultan al país interesado para contactar con las autoridades del país en cuya base de datos se halla

[659] No obstante, el mencionado precepto añade: "*Si no pueden utilizarse las impresiones dactilares de dicha persona o la búsqueda con las impresiones dactilares es infructuosa, la búsqueda se llevará a cabo mediante los datos a que se refiere el artículo 9, apartado 4, letras a) o c). Dicha búsqueda podrá llevarse a cabo en combinación con los datos a que se refiere el artículo 9, apartado 4, letra b).*"

la huella dactilar que ha hecho *match* con la introducida, y averiguar así la identidad del sujeto al que pertenece.

d.2) ***Posibles utilidades en la instrucción de las causas***

Tal y como se ha expuesto, si bien las técnicas de identificación por huellas dactilares no son nuevas, lo cierto es que su combinación con la IA sí lo es y puede resultar significativamente útil en el ámbito de la investigación criminal, habida cuenta de su elevado índice de precisión, su aceptable precio de implementación (especialmente si se tiene en cuenta la relación coste-beneficio) y su sencillo funcionamiento.

d.2.a) Incremento de la eficacia en la búsqueda e identificación de presuntos delincuentes, víctimas, testigos, cadáveres y personas desaparecidas

El hecho de que los sistemas de IA de reconocimiento de huellas dactilares cuenten con vastas bases de datos interconectadas a nivel nacional e internacional y dispongan de unos algoritmos entrenados para examinar millones de imágenes dactilares en tiempo récord, extraer patrones y hallar coincidencias, facilita sin duda la tarea de los cuerpos de policía y de las autoridades fiscales y judiciales a la hora de proceder a identificar (*ab initio* o a modo de confirmación) a aquellos individuos que puedan resultar de interés para la investigación criminal.

Y es que imaginemos que se perpretra un ataque terrorista en un aeropuerto español y se hallan restos de huellas dactilares que podrían servir para identificar al presunto autor de los hechos. Pensemos, asimismo, que este es un ciudadano holandés que ya había sido condenado en su país por sendos delitos graves y, por ende, las imágenes de sus huellas dactilares se hallan recogidas en su base de datos nacional y, asimismo, están introducidas en el sistema SAID. Imaginemos, también, que resulta muy urgente prodecer a la identificación del mencionado terrorista habida cuenta de que esta puede llevar a averiguar la identidad de otras personas implicadas y, sobre todo, puede ayudar a prevenir la comisión de más actos terroristas planificados para ser llevados a cabo en las próximas horas.

En tal caso, de no existir un sistema de IA potente que, mediante la introducción de las imágenes dactilares obtenidas, analizara en cuestión de minutos toda la información contenida en las bases de datos disponibles, por ejemplo, y procediera a su filtración, aportando así a la policía española de forma rápida uno o varios candidatos (en función de los *match* obtenidos) para que esta pudiera hacer una evaluación final y tomar decisiones, tal tarea hubiera resultado larga (pudiendo durar días o incluso semanas, habida cuenta de la necesidad de consultar con bases de datos de otros países, a ciegas) y seguramente, ineficaz para los fines buscados, a saber, detener a los presuntos autores de los hechos y evitar la perpretración de más actos terroristas por parte del mismo comando.

No obstante, bajo mi punto de vista, en aquellos casos en que se investiguen delitos graves, puede resultar insuficiente el acceso a las bases de datos policiales, siendo que además podría resultar idóneo y necesario acceder a aquellas bases de datos de huellas dactilares del DNI y pasaporte de los ciudadanos, en este caso, de la UE. Y es que en la actualidad, como se ha dicho, las huellas dactilares que nos toman para plasmar en tales documentos oficiales son únicamente empleadas con fines de identificación oficial y lo único que podría llevarse a cabo sería la comparación concreta de una muestra dactilar de un sospechoso con la que este cedió para la realización de su DNI o pasaporte, con autorización judicial, pero ello resulta muy limitado.

Y es que, imaginemos que en el caso anteriormente expuesto se hallan asimismo huellas dactilares que no corresponden al mencionado ciudadano holandés, se introducen en las bases de datos policiales a las que las autoridades tienen acceso y no se produce *match* o coincidencia alguna. Ante ello, sin duda, haría falta un paso más y, en mi opinión, la causa lo merece. Así, si mediante autorización judicial, previo informe del Ministerio Fiscal, pudiera accederse a analizar de forma automática las imágenes biometrizadas de las huellas dactilares de los ciudadanos europeos cedidas para la realización del pasaporte o DNI (previa información a estos de que sus huellas dactilares podrían ser utilizadas a efectos de investigación criminal en casos de delitos graves -legalmente previstos- o búsqueda de personas desaparecidas, previa autorización judicial), el abanico de posibilidades de investigación se abriría más que significativamente.

Y es importante poner de manifiesto, además, que ello no solo podría resultar eficaz y necesario para identificar a los presuntos autores de los hechos, que en múltiples ocasiones ya se encuentran "fichados" por la policía, sino que podría resultar un avance enorme para identificar víctimas, ya que sus huellas dactilares no suelen constar en las bases de datos policiales.

Así, pensemos en las víctimas del ya expuesto caso del ataque terrorista en un aeropuerto español. Pensemos, por un lado, en los momentos de caos e incertidumbre que, sin duda, vivirían las familias de aquellas personas que supuestamente estaban en dicho recinto en el momento de los hechos; y, por otro lado, en la sensación de impotencia que invadiría a los agentes de policía que carecieran de noticias sobre varios de los cuerpos hallados sin vida. Y es que en ambos casos resultaría absolutamente idóneo y pertinente, bajo mi punto de vista, proceder rápida y eficazmente a cotejar las muestras dactilares obtenidas con aquellas contenidas en las bases de datos policiales, nacionales y europeas y, en su defecto, y previa autorización judicial, en las bases de datos de identificación oficial (DNI y pasaporte) asimismo nacionales y europeas, a los efectos de, por un lado, confirmar la identidad de los supuestamente fallecidos y, por otro lado, obtener la identidad de aquellos indocumentados carentes de otra información identificativa útil.

Y es que no considero que ello pueda vulnerar ningún derecho fundamental de las personas afectadas, habida cuenta de que se llevaría a cabo siempre mediante autorización judicial y únicamente en caso de concurrencia de aquellos delitos graves legalmente previstos (según la voluntad del legislador de cada momento), así como en casos de necesidad de identificación de cadáveres y personas desaparecidas.

d.2.b) Extracción de huellas dactilares de fotografías y ulterior identificación

Investigadores del Instituto Nacional de Informática de Japón (NII) han extraído con éxito huellas dactilares de fotografías (tomadas a una distancia de hasta tres metros) a través de las imágenes de los dedos expuestos por los usuarios. En concreto, el experimento se llevó

a cabo con fotografías tomadas por jóvenes haciendo el signo de la paz con sus dos dedos mirando a la cámara.[660]

Ello, desde luego, puede resultar de gran utilidad para la investigación de las causas, especialmente para la identificación de personas de interés en aquellos casos en que se disponga de fotografías que muestren las huellas dactilares y estas sean lo suficientemente nítidas como para poder extraer los datos biométricos y proceder a su análisis.

e) Reconocimiento de ADN

e.1) *Concepto*

A modo de concepto puede decirse que el reconocimiento de ADN (Ácido Desoxirribonucleico) es aquella tecnología que permite, a través de la IA, la identificación de personas y/o la comprobación de su identidad mediante el análisis y, en su caso, la comparación de sus datos biométricos, en concreto, los genomas humanos, mediante la búsqueda de coincidencias o similitudes entre millones de códigos genéticos.

Si bien las técnicas de análisis de ADN tienen múltiples utilidades, especialmente médicas, desde hace ya varias décadas[661] estas (con el método tradicional) vienen siendo empleadas en el ámbito de la investigación criminal, habiéndose convertido en la prueba estrella para identificar (y descartar, en su caso) tanto la participación de delincuentes como de víctimas, dependiendo del caso de que se trate. Y es que a día de hoy el grado de precisión de los análisis forenses de ADN (llevados a cabo en España, principalmente, por los seis laboratorios con los que cuenta la Policía Nacional, el laboratorio con el que cuenta la Guardia Civil, el laboratorio del Instituto Nacional de Toxicología y Ciencias Forenses y los laboratorios de Mossos d'Esquadra, Ertzaintza y Guardia Foral de Navarra, así como los laboratorios acreditados por la Comisión Nacional para el uso forense del ADN) es ciertamente elevadísimo, ya que en la mayoría de los casos, tal y como he podido observar por mi experiencia, alcanza el 99,9% de fiabilidad.

[660] Véase Perala, 2017.
[661] En concreto, desde 1988, año en que, por primera vez, en Reino Unido, el ADN fue empleado para identificar y condenar a un delincuente.

Así, actualmente, las mencionadas técnicas permiten, por ejemplo, cotejar el perfil genético de un sospechoso (muestra genética indubitada) con el contenido genético hallado en el lugar del crimen (muestra genética dubitada) para confirmar o descartar su participación en el mismo, erigiéndose como una prueba clave tanto de cargo como de descargo; averiguar la identidad ignorada de un delincuente introduciendo el material genético recogido en el lugar de los hechos (muestra genética dubitada) en una gran base de datos para verificar si tiene correspondencia o no con alguno de los contenidos en la misma (muestras genéticas indubitadas); comprobar la identidad de una víctima hallada fallecida o en estado de inconsciencia, indocumentada, mediante la introducción de su información genética (muestra genética dubitada) en una gran base de datos para detectar coincidencias (muestras genéticas indubitadas); o comprobar y verificar la identidad de una presunta víctima irreconocible mediante la comparación de su material genético (muestra genética dubitada) con la información genética aportada por la familia (muestras genéticas indubitadas), entre otras.

En España el uso del ADN para fines de investigación criminal viene regulado, principalmente, por la LO 10/2007, de 8 de octubre, reguladora de la base de datos policial sobre identificadores obtenidos a partir del ADN (en adelante, LO 10/2007); el Real Decreto 1977/2008, de 28 de noviembre, por el que se regula la composición y funciones de la Comisión Nacional para el uso forense del ADN (en adelante, RD 1977/2008); la LECrim; el Código Penal; y la normativa europea e internacional.

Por un lado, en virtud de la LO 10/2007, tal y como se desprende de su Exposición de Motivos, existe en nuestro país una base de datos en la que se integran los ficheros de los Cuerpos y Fuerzas de Seguridad del Estado que almacenan "*los datos identificativos obtenidos a partir de los análisis de ADN que se hayan realizado en el marco de una investigación criminal, o en los procedimientos de identificación de cadáveres o de averiguación de personas desaparecidas*", siendo posible que "*los resultados obtenidos a partir del análisis de las muestras biológicas del sospechoso, detenido o investigado, sean inscritos y conservados en la base de datos policial, a fin de que puedan ser utilizados en esa concreta investigación, o en otras que se sigan por la comisión de alguno de los delitos*

para los que la propia Ley habilita la inscripción de los perfiles de ADN en la base de datos."

En la mencionada base de datos, no obstante, únicamente puede incluirse información genética no codificante -es decir, reveladora exclusivamente de la identidad de la persona y de su sexo- obtenida a partir del ADN en el marco de una investigación criminal. Y es que en concreto, en virtud de lo dispuesto en el artículo 3.1 de la antedicha ley, deben inscribirse en dicha base de datos:

-los datos identificativos obtenidos a partir del ADN de un individuo en aquellos casos en que este haya prestado su consentimiento de forma expresa; -
y, asimismo, en aquellos casos en que, aun no constando consentimiento expreso del afectado, se trate de:

a) datos identificativos obtenidos a partir del ADN "*de muestras o fluidos que, en el marco de una investigación criminal, hubieran sido hallados u obtenidos a partir del análisis de las muestras biológicas del sospechoso, detenido o investigado, cuando se trate de delitos graves y, en todo caso, los que afecten a la vida, la libertad, la indemnidad o la libertad sexual, la integridad de las personas, el patrimonio siempre que fuesen realizados con fuerza en las cosas, o violencia o intimidación en las personas, así como en los casos de la delincuencia organizada, debiendo entenderse incluida, en todo caso, en el término delincuencia organizada la recogida en el artículo 282 bis, apartado 4 de la LECrim en relación con los delitos enumerados*."; o

b) patrones identificativos recogidos en los procedimientos de identificación de personas desaparecidas o de restos cadavéricos.

En virtud de lo dispuesto en el artículo 7.3 de la mencionada ley, los datos contenidos en dicha base de datos (que deberán ser cancelados y actualizados conforme a lo previsto en el artículo 9) pueden ser cedidos, para la investigación de delitos, tanto a las autoridades judiciales, fiscales o policiales de aquellos países que tengan convenio internacional vigente ratificado por España en tal sentido; como a las Policías Autonómicas, que, no obstante, solo podrán utilizar tal información para la investigación de los delitos

enumerados en el artículo 3.1.a) o, en su caso, para la identificación de cadáveres o de personas desaparecidas; y al Centro Nacional de Inteligencia, que podrá utilizar dichos datos para el cumplimiento de sus funciones relativas a la prevención de tales delitos.

Por otro lado, RD 1977/08 regula la denominada Comisión Nacional para el uso forense del ADN, como órgano colegiado adscrito al Ministerio de Justicia, dependiente jerárquicamente de la Secretaría de Estado de Justicia, que tiene como funciones, en virtud de lo dispuesto en su artículo 3, principalmente, la acreditación de los laboratorios facultados para contrastar perfiles genéticos en la investigación y persecución de delitos y la identificación de cadáveres o averiguación de personas desaparecidas; la evaluación de su cumplimiento y el establecemiento de los controles oficiales de calidad a los que estos deban someterse de forma periódica; la fijación de criterios de coordinación entre los antedichos laboratorios y su funcionamiento, así como el análisis de todos aquellos aspectos científicos y técnicos, organizativos, éticos y legales que garanticen el buen funcionamiento de los mismos; y la determinación de las condiciones de seguridad en la custodia y la fijación de todas las medidas que garanticen la estricta confidencialidad y reserva de las muestras, los análisis y los datos que se obtengan.

Especial mención merece también el Real Decreto 1110/2015, de 11 de diciembre, por el que se crea y regula el Registro Central de Delincuentes sexuales, que contiene toda la información penal que consta tanto en el Registro Central de Penados como en el Registro Central de Sentencias de Responsabilidad Penal de los Menores respecto de quienes hayan sido condenados en sentencia firme (tanto en España como en otros países, en particular en los Estados miembros de la Unión Europea y del Consejo de Europa) por cualquier delito contra la libertad e indemnidad sexuales, así como por trata de seres humanos con fines de explotación sexual (inclusive la pornografía infantil), incluyendo el código identificador del perfil genético (ADN) de los condenados cuando así se haya acordado por el órgano judicial.

Por su parte, en la LECrim, el artículo 326 dispone, respecto de la recogida de vestigios cuyo análisis genético pueda contribuir al esclarecimiento de los hechos investigados (muestras biológicas dubitadas), que esta puede llevarse a cabo o bien directamente por la Policía Judicial, en aquellos casos en que hubiera peligro de desaparición, o bien a través

de autorización judicial, en el resto de casos. Así: "*Cuando se pusiera de manifiesto la existencia de huellas o vestigios cuyo análisis biológico pudiera contribuir al esclarecimiento del hecho investigado, el Juez de Instrucción adoptará u ordenará a la Policía Judicial o al médico forense que adopte las medidas necesarias para que la recogida, custodia y examen de aquellas muestras se verifique en condiciones que garanticen su autenticidad*", sin perjuicio de lo establecido en el artículo 282, que determina que "*La Policía Judicial tiene por objeto y será obligación de todos los que la componen, averiguar los delitos públicos que se cometieren en su territorio o demarcación; practicar, según sus atribuciones, las diligencias necesarias para comprobarlos y descubrir a los delincuentes, y recoger todos los efectos, instrumentos o pruebas del delito de cuya desaparición hubiere peligro, poniéndolos a disposición de la autoridad judicial*."

El artículo 363 LECrim, respecto de la recogida y el análisis químico de muestras indubitadas de los sospechosos que permitan determinar, por comparación y verificación, la identidad de vestigios genéticos dubitados, establece la necesidad de autorización y control judicial en aquellos casos en los que estos no presten su consentimiento de forma expresa. Así:

"*Los Juzgados y Tribunales ordenarán la práctica de los análisis químicos únicamente en los casos en que se consideren absolutamente indispensables para la necesaria investigación judicial y la recta administración de justicia.*

Siempre que concurran acreditadas razones que lo justifiquen, el Juez de Instrucción podrá acordar, en resolución motivada, la obtención de muestras biológicas del sospechoso que resulten indispensables para la determinación de su perfil de ADN. A tal fin, podrá decidir la práctica de aquellos actos de inspección, reconocimiento o intervención corporal que resulten adecuados a los principios de proporcionalidad y razonabilidad."

Y, por su parte el artículo 520.6.c) LECrim, al referirse a las labores de asistencia del Letrado del detenido, dispone que una de ellas es la de informarle de las consecuencias de la prestación o denegación de consentimiento para la práctica de diligencias que se le

soliciten y, en concreto, establece que "*Si el detenido se opusiera a la recogida de las muestras mediante frotis bucal, conforme a las previsiones de la Ley Orgánica 10/2007, de 8 de octubre, reguladora de la base de datos policial sobre identificadores obtenidos a partir del ADN, el juez de instrucción, a instancia de la Policía Judicial o del Ministerio Fiscal, podrá imponer la ejecución forzosa de tal diligencia mediante el recurso a las medidas coactivas mínimas indispensables, que deberán ser proporcionadas a las circunstancias del caso y respetuosas con su dignidad*."

Además, en el Código Penal, el artículo 129 bis faculta a los jueces y tribunales a acordar la toma de muestras biológicas y su análisis genético como consecuencia accesoria de la sentencia de condena en aquellos casos en que se hayan cometido ciertos delitos: "*Si se trata de condenados por la comisión de un delito grave contra la vida, la integridad de las personas, la libertad, la libertad o indemnidad sexual, de terrorismo, o cualquier otro delito grave que conlleve un riesgo grave para la vida, la salud o la integridad física de las personas, cuando de las circunstancias del hecho, antecedentes, valoración de su personalidad, o de otra información disponible pueda valorarse que existe un peligro relevante de reiteración delictiva, el juez o tribunal podrá acordar la toma de muestras biológicas de su persona y la realización de análisis para la obtención de identificadores de ADN e inscripción de los mismos en la base de datos policial. Únicamente podrán llevarse a cabo los análisis necesarios para obtener los identificadores que proporcionen, exclusivamente, información genética reveladora de la identidad de la persona y de su sexo. Si el afectado se opusiera a la recogida de las muestras, podrá imponerse su ejecución forzosa mediante el recurso a las medidas coactivas mínimas indispensables para su ejecución, que deberán ser en todo caso proporcionadas a las circunstancias del caso y respetuosas con su dignidad*."

Respecto de la normativa europea, procede hacer especial alusión a la ya anteriormente mencionada "Decisión Prüm". Y es que en virtud de tal Decisión, los Estados Miembros de la UE tienen la obligación de crear bases de datos nacionales de ADN (al perfil genético se le da un código alfanumérico de referencia que, no obstante, no permite la identificación de su titular de forma directa) que pueden ser consultadas por los cuerpos policiales del resto de países para fines de investigación criminal. Así, por un lado, cada país miembro actualiza a diario los códigos de referencia de los distintos perfiles de ADN de que dispone

en una base de datos local (la de cada cuerpo policial), que pasa directamente a la base de datos nacional y a la base de datos europea. Por otro lado, si las autoridades policiales de alguno de tales países miembros, en el marco de una investigación criminal, necesitan realizar una consulta en la mencionada base de datos europea, pueden introducir el perfil genético dubitado para ver si tiene concordancia o no con alguno de los perfiles genéticos indubitados contenidos en la misma, y en caso de que exista coincidencia o *match*, pueden solicitar al país de donde procede la muestra con interés criminal que les proporcione datos sobre la identidad de su titular.

Especial mención debe hacerse también de lo dispuesto en relación al uso del ADN para los fines del sistema SIS en el Reglamento (UE) 2018/1862 del Parlamento Europeo y del Consejo, de 28 de noviembre de 2018, relativo al establecimiento, funcionamiento y utilización del SIS en el ámbito de la cooperación policial y de la cooperación judicial en materia penal, que en su Considerando 26 dispone: "*En casos claramente definidos en que no se disponga de datos dactiloscópicos, ha de ser posible añadir a la descripción un perfil de ADN. Únicamente los usuarios autorizados deben tener acceso a ese perfil de ADN. Los perfiles de ADN deberían facilitar la identificación de personas desaparecidas que necesitan protección y especialmente de menores desaparecidos, en particular si se permite el uso de los perfiles de ADN de sus ascendientes o descendientes directos o hermanos para permitir la identificación. Los datos de ADN han de contener únicamente la información mínima necesaria para la identificación de la persona desaparecida*." y en su Considerando 27 establece: "*En el SIS solo deben consultarse perfiles de ADN cuando la identificación sea necesaria y proporcionada a efectos de lo dispuesto en el presente Reglamento. No se deben consultar ni tratar perfiles de ADN para fines distintos de aquellos para los que fueron introducidos en el SIS. Deben aplicarse las normas sobre seguridad y protección de datos establecidas en el presente Reglamento. Se deben tomar, si fuera necesario, medidas adicionales cuando se utilicen perfiles de ADN con el fin de evitar todo riesgo de falsas coincidencias, piratería informática o intercambio no autorizado con terceros*." Por su parte, el artículo 42, que fija normas específicas aplicables a la introducción de fotografías imágenes faciales, datos dactiloscópicos y perfiles de ADN, en su apartado 3 dispone el uso subsidiario y restrictivo de estos últimos: "*Solo se podrá añadir a las descripciones un perfil de ADN en las situaciones previstas en el artículo 32,*

apartado 1, letra a)[662]*, únicamente tras un control de calidad que determine si se cumplen las normas mínimas de calidad de los datos y las especificaciones técnicas y solo cuando no se disponga de fotografías, imágenes faciales o datos dactiloscópicos o estos no sean adecuados para la identificación. Los perfiles de ADN de personas que sean ascendientes o descendientes directos o hermanos de la persona objeto de la descripción podrán añadirse a la descripción siempre que dichas personas den su consentimiento expreso. Cuando se añada un perfil de ADN a una descripción, dicho perfil deberá contener la mínima información estrictamente necesaria para la identificación de la persona desaparecida*."

Y, en el ámbito internacional, con aquellos países a los que no les es de aplicación la "Decisión Prüm", la búsqueda automatizada de coincidencias entre perfiles genéticos se lleva a cabo a través de Interpol, que hace una labor de intermediaria (siempre y cuando no exista un espefícico Tratado bilateral o multilateral entre España y otro/s país/es que determine otra forma de proceder). Así, en caso de que las autoridades policiales de nuestro país tuvieran interés en cotejar una muestra dubitada de ADN con las muestras de ADN indubitadas contenidas en la base de datos de Marruecos, por ejemplo, deberían remitir tal perfil genético a Interpol y solicitar su cotejo con la base de datos de tal país. Posteriormente, dicha organización internacional sería la que remitiría el resultado, positivo o negativo, a España.

En relación con ello, es importante hacer referencia al sistema "Combined DNA Index System" (CODIS), el programa informático policial del FBI, cedido a múltiples cuerpos policiales de todo el mundo de forma gratuita, que hace posible que las autoridades puedan intercambiar y comparar perfiles de ADN de forma automática a partir de la interconexión de las bases de datos genéticos, lo cual resulta un avance sin precedentes.

Una vez analizada la regulación relativa a las técnicas de análisis automático de muestras biológicas en el ámbito de la investigación criminal existentes en la actualidad, al alcance

[662] Es decir, en lo relativo a personas desaparecidas que requieran medidas de protección para su propia protección o para prevenir una amenaza para el orden público o la seguridad pública.

de nuestro país, procede hacer la siguiente reflexión, compartida por expertos en análisis de ADN de distintos cuerpos policiales con los que he tenido la oportunidad de conversar.

Y es que con los sistemas de búsqueda y comparación automática con los que actualmente se cuenta, que además cumplen con los estándares exigidos por la LO 7/21, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, y el RGPD, ya existen unos sistemas lo suficientemente sólidos y eficientes que no justifican la necesidad de pensar en ir más allá, al menos de forma inmediata, siendo que cuando algo funciona, y funciona bien, es mejor no realizar cambios que puedan hacer tambalear la eficacia de sus exitosos resultados (siempre sin olvidar, claro está, el interés por la evolución que, no obstante, no debe ofuscarnos en exceso). Así, en la actualidad, en el ámbito de la investigación criminal, el único cambio que podría introducirse respecto del análisis automático de ADN sería el hecho de que, una vez producido el *match* entre dos perfiles (uno dubitado y otro indubitado), se intercambiara de forma directa y mecánica la información sobre la identidad del titular, pero las autoridades policiales (en la misma línea que la normativa sobre protección de datos) entienden que la mejor opción para proteger el derecho a la privacidad y a la protección de datos de los ciudadanos es llevar a cabo en tales casos un contacto personal entre autoridades para tratar con dichas informaciones, lo cual no me parece mal, ya que asegura y refuerza el buen resguardo de datos extremadamente sensibles.

No obstante, cierto es que el análisis de ADN es la técnica de identificación biométrica más cara de que en la actualidad disponen las autoridades policiales y judiciales y, si bien hasta ahora el coste se imputaba principalmente a los medios materiales y personales de los laboratorios especializados, hoy en día hay que sumar también el precio de los potentes y sofisticados sistemas que se requieren para llevar a cabo las tareas de cotejo automático que, además, no eliminan (aunque sí disminuyen), al menos por el momento, la necesidad de contar con personal humano experto para llevar a cabo las tareas de recogida de muestras, introducción en bases de datos, contacto entre autoridades, análisis de resultados y comprobación del veredicto final, por lo que tampoco es la panacea.

e.2) ***Posibles utilidades en la instrucción de las causas***

e.2.a) Incremento de la eficacia en la búsqueda e identificación de presuntos delincuentes, víctimas, testigos, cadáveres y personas desaparecidas

Al igual que se ha puesto de manifiesto respecto de los sistemas de IA de reonocimiento automático de huellas dactilares, el hecho de que los sistemas de IA de reconocimiento automático de restos biológicos cuenten con vastas bases de datos interconectadas a nivel nacional e internacional y dispongan de unos algoritmos entrenados para examinar millones de secuencias de ADN en tiempo récord, extraer patrones y hallar coincidencias, facilita sin duda la tarea de los cuerpos de policía y de las autoridades fiscales y judiciales a la hora de proceder a identificar (*ab initio* o a modo de confirmación) a aquellos individuos que puedan resultar de interés para la investigación criminal.

Y es que, por un lado, imaginemos que se perpretra un robo armado, con detención ilegal, en un domicilio por parte de una organización criminal extranjera que únicamente está en España, de paso, con el fin de cometer cuantos más delitos contra el patrimonio como le sea posible. Pensemos, también, que en el lugar de los hechos se hallan restos biológicos de uno de los asaltantes, habida cuenta de que se dejó un guante, y que este ya es un viejo conocido de la policía francesa, por ejemplo, siendo que en tal país ha sido condenado en varias ocasiones por delitos graves, razón por la cual su ADN se halla registrado en la base de datos nacional accesible por los Estados Miembros. Imaginemos, asimismo, que no existe otra fuente de información sobre la identidad de dicho delincuente, siendo que es la primera vez que opera en nuestro país, y sin embargo resulta necesario dar con su identidad para truncar su plan y poder prevenir así la comisión de más delitos.

Así, mediante el uso de un sistema de IA de reconocimiento automático de ADN, en el caso expuesto podrían introducirse los restos biológicos hallados para ser comparados con los contenidos en las vastas bases de datos accesibles a las autoridades y dar, así, con la identidad del delincuente de forma mucho más rápida y efectiva que en caso de hacerlo de forma "manual", siendo que especialmente en este supuesto, habida cuenta del componente internacional, las tareas serían mucho más difíciles y lentas de lo normal.

Por otro lado, por ejemplo, imaginemos que se halla por la policía un cadáver en avanzado estado de descomposición y lo único que puede tener valor a efectos identificativos son las los restos biológicos de este. Pensemos, asimismo, que estos se introducen en la base de

datos de ADN de personas desaparecidas y el sistema de IA detecta, de forma rápida y precisa, una coincidencia con una secuencia de ADN de un ciudadano austriaco desaparecido, permitiendo así resolver un caso que llevaba meses lleno de incógnitas, lo cual no hubiera sido posible en caso de emplear métodos más tradicionales.

Y es que vemos que, sin duda, la automatización y la introducción de IA en estos supuestos implica un significativo avance en las tareas de identificación y, por ende, en la investigación criminal.

e.2.b) Facilitación de las tareas de identificación en los casos más complejos

Los sistemas de reconocimiento genético a través de IA pueden ayudar a marcar la diferencia y ser de gran utilidad para aquellas tareas más complejas. Y es que en ocasiones, ciertos análisis de ADN requieren de un esfuerzo adicional (que para un humano resulta muy difícil de realizar) que podría ser llevado a cabo por "máquinas" que funcionen con técnicas de aprendizaje automático muy potentes.

Así, por ejemplo, en aquellos casos en que dos o más personas hayan estado en contacto con un mismo objeto o persona, como podría ocurrir en una agresión sexual múltiple, los perfiles genéticos que se obtienen tienden a estar mezclados y dan lugar a arduas tareas de análisis, puesto que las combinaciones numéricas que resultan son elevadas, lo que dificulta mucho la posibilidad de determinar el número de individuos que hay detrás de esa información genética y separar y diferenciar los códigos genéticos de todos ellos para lograr identificarlos. Ello, sin embargo, puede ser realizado de forma más sencilla con la ayuda de la IA, tal y como pensaron los investigadores de la Universidad de Syracuse (NY, EEUU) que, tras años de trabajo, con el respaldo del National Institute of Justice (NIJ) estadounidense, han creado el *software* PACE ("*Probabilistic Assessment for Contributor Estimation*"), cuya licencia ha sido conferida a la compañía norteamericana NicheVisions Forensics, LLC., que presenta una herramienta de aprendizaje automático que permite una gestión de perfiles genéticos rápida y precisa, capaz de predecir en cuestión de segundos el número de contribuyentes existentes en mezclas complejas de ADN, lo que sin duda

resulta de gran ayuda para los analistas humanos a fin de llegar a conclusiones más fiables y precisas.[663]

e.2.c) Reconstrucción de rostros humanos en 3D para su análisis mediante técnicas de reconocimiento facial

Asímismo interesante es hacer referencia a las novedosas técnicas de IA que permiten la reconstrucción de rostros humanos en 3 dimensiones (3D) a partir del ADN.

En relación con ello, hay que mencionar el *software* desarrollado en 2015 por un equipo de investigación de la Penn State University (Pensilvania, EEUU), dirigido por el profesor de antropología y genética Mark Shriver, capaz de crear, de forma predictiva, un modelo tridimensional (3D) de una cara humana a partir de una muestra de ADN. Y es que la mencionada herramienta de IA lleva a cabo la mencionada representación en cuestión de minutos mediante la traza de conexiones entre marcadores genéticos y puntos del rostro,[664] lo que sin duda supone un gran avance para la investigación criminal, puesto que no solo permite cotejar las muestras de ADN halladas en el lugar de los hechos con las contenidas en las bases de datos policiales, sino que además hace posible introducir el retrato robot realizado en las bases de datos de imágenes para ver si es identificado por un sistema de reconocimiento facial y, asimismo, puede ser sometido a reconocimiento por víctimas y testigos.

Así, imaginemos que en un caso de agresión sexual la víctima acude a la policía y manifiesta no tener conocimiento de la identidad del agresor, que huyó del lugar de los hechos, y asegura asimismo que no puede dar datos sobre sus rasgos físicos puesto que iba tapado con un pasamontañas, gafas de sol y ropa gruesa. Pensemos, no obstante, que pudieron ser recogidas por el médico forense muestras de ADN inmediatamente después de la agresión, cuando la víctima acudió al hospital. Imaginemos, asimismo, que gracias a la tecnología expuesta se recrea la imagen facial del presunto autor de los hechos, mediante la creación de un retrato robot, que al ser introducido en la base de datos de imágenes de delincuentes sexuales habituales resulta coincidir con uno de de estos, lo que lleva a la

[663] Véase Marciano & Adelman, 2017.
[664] Véase Murphy, 2015.

policía a identificar al mencionado delincuente y a hacer un cotejo entre su ADN y el hallado en el lugar de los hechos, previa autorización judicial. Y, finalmente, pensemos que el resultado que arroja tal diligencia de investigación es positivo, con un grado de fiabilidad superior al 99%, lo que no deja dudas sobre la identidad del autor.

No obstante, y a pesar de la enorme utilidad que aparentemente pueden tener este tipo de sistemas de IA, lo cierto es que existe mucho escepticismo al respecto. Y es que, por un lado, es una tecnología todavía muy en ciernes, por lo que aun es vista como "ciencia ficción" por la mayoría de expertos de todo el mundo; y, por otro lado, cierto es que un perfil genético puede dar datos sobre el rostro que en vez de conducir a la identificación de una persona, la dificulten. Así, si el ADN manifiesta que una persona tiene pelo castaño, ojos azules y rostro angular, y el *software* recrea un rostro con tales características, las investigaciones irán dirigidas a hallar a alguien con tal apariencia, pero en realidad, quizás tal indidivuo en la actualidad tiene el pelo rapado, lleva lentillas de otro color, y cuenta con sobrepeso, por lo que la imagen representada poco o nada se corresponderá con la realidad, lo que debe llevar a emplear tal tipo de herramientas con una enorme cautela.

e.2.d) Incremento de la eficacia en la búsqueda e identificación de objetos mediante el uso de ADN sintético/artificial

Si bien todavía su uso no está muy extendido, en la actualidad ya resulta posible emplear ADN sintético, con un código único que suele resultar visible únicamente a través de luz ultravioleta, que se secuencia en un laboratorio y se introduce en una base de datos, con el fin de marcar objetos y poder así localizarlos en caso de pérdida y, principalmente, sustracción.

Y es que son ya varias las empresas que en distintos países del mundo se han dedicado a crear secuencias de ADN sintético con fines de investigación criminal, ya que su uso es, por un lado, disuasorio, y por otro lado, útil, especialmente para la recuperación de objetos sustraídos, como por ejemplo joyas.

Así, al igual que un teléfono móvil robado (previa denuncia) puede recuperarse por su propietario, en caso de ser hallado, mediante la comprobación de su número de IMEI, hoy

en día puede procederse a marcar objetos personales de modo que, en caso de sustracción y posterior hallazgo, pueda ser introducido su ADN sintético en una base de datos y dar con su propietario de forma rápida y precisa.

Y es que pensemos en un caso de robo de múltiples piezas en una joyería en Madrid. En circunstancias normales, los hechos serían denunciados por su propietario a la policía, mediante la aportación de fotografías y facturas, se procedería a su tasación pericial y el caso quedaría al albor de que el autor del robo (o un tercero) fuera hallado con el material sustraído y este fuera identificado por la policía como propio del joyero denunciante. Ello, no obstante, es tarea muy compleja en caso de que las joyas se hallen al cabo del tiempo en un lugar distinto de aquel donde ocurrió el robo, puesto que resulta imposible realizar una comprobación, una a una, de todas las imágenes de joyas sustraídas que constan en los distintos cuerpos policiales, por lo que con toda probabilidad, el dueño nunca recuperaría su mercancía. No obstante, en caso de que tales joyas, antes de ser sustraídas, hubieran sido marcadas con ADN sintético o artificial, las tareas de investigación se verían facilitadas de forma radical, y lo mismo ocurriría con la recuperación de los objetos robados, habida cuenta de que si estos fueran hallados, podrían enseguida ser entregados a su propietario.

Y tal técnica, ya empleada desde hace más de una década[665] (entre otras, para marcar cables de cobre ferroviarios, que suelen ser objeto de sustracción de modo recurrente) si se combina con la IA, aumentaría sin duda exponencialmente su capacidad de búsqueda e identificación, habida cuenta de que la automatización de tales funciones supondría un avance cualitativo sin precedentes.

f) Reconocimiento de firma y de escritura

f.1) *Concepto*

A modo de concepto, puede decirse que el reconocimiento de firma y escritura es aquella tecnología que permite, a través de la IA, la identificación de personas y/o la comprobación de su identidad mediante el análisis y, en su caso, la comparación de sus datos biométricos,

[665] Véase La Vanguardia, 2011.

a saber, en este caso, signos y símbolos manuscritos plasmados en un soporte físico o digital.

Si bien a lo largo de los siglos la firma y la escritura han sido uno de los elementos de identificación y verificación de la identidad más empleados en todo el mundo, lo cierto es que en la actualidad, y desde hace ya varias décadas, sobre todo a partir de la irrupción de los sistemas informáticos, estos han ido perdiendo valor, ya que cada vez se van dejando más atrás tanto la rúbrica como la caligrafía manuscrita, que van dando paso a la firma digital y a la escritura automática.

No obstante, gracias a la aparición de dispositivos electrónicos e informáticos (teléfonos móviles, tabletas, etc) que incluyen la posibilidad de interacción mediante interfaces táctiles, como lápices o punteros, el potencial identificador de la firma y la escritura manuscritas está poco a poco recobrando fuerza. Además, si bien en ocasiones la firma digital se lleva a cabo a través de un chip habilitado que simplemente deja rastro de nuestro nombre, apellidos y hora de la misma, la mayoría de veces esta consiste en una mera rúbrica manuscrita digitalizada, lo que permite extraer patrones singulares para su posterior análisis.

Y es que, en relación con ello, es importante poner de manifiesto que no es lo mismo la firma digital, que es un archivo informático que no contiene dato biométrico alguno, y la firma digitalizada, que es la rúbrica manuscrita de una persona realizada en un soporte que permite su digitalización. No obstante, respecto de esto último, procede decir que si el mencionado soporte no cuenta con una tecnología que detecte los datos biométricos (imaginemos que simplemente sirve para pasar la firma manuscrita a formato PDF), la tarea de reconocimiento posterior resultará más compleja, habida cuenta de que características tales como la presión o la velocidad del trazo no podrán ya tenerse en cuenta.

Hasta el momento, bien es sabido que en el ámbito de la investigación criminal el análisis y reconocimiento de la firma y escritura humanas se llevaba a cabo, especialmente, por expertos peritos calígrafos que, principalmente mediante técnicas tradicionales de cotejo estático (consistentes en la verificación del parecido entre dos firmas y caligrafías), y de

prueba dinámica (consistentes en el estudio de la forma y la velocidad de estas), sacaban conclusiones, no siempre del todo concluyentes.

No obstante, en la actualidad, ya existen sistemas que, a través de la IA, realizan las mencionadas funciones de forma automática.

Así, básicamente, se lleva a cabo o bien una tarea de identificación, introduciendo en una base de datos la firma o la escritura de que se disponga (muestra dubitada), a los efectos de averiguar si se corresponden con las de alguno de los individuos que constan registrados (muestras indubitadas); o bien una tarea de verificación/autenticación de identidad, comparando la firma o la escritura de que se disponga (muestra dubitada) con la de la persona que presuntamente las ha realizado (muestra indubitada).

En el ámbito de la firma digitalizada, por ejemplo, principalmente las entidades financieras ya emplean sistemas de IA de identificación automatizada de sus clientes con el ánimo de evitar fraudes. De tal modo, cada vez que uno de sus clientes plasma su firma en un soporte digitalizado (normalmente, una tableta), el sistema la compara con la firma que el banco ya tiene de él en su base de datos, de forma que bloquea la operación en caso de detectar que no existe identidad entre ambas rúbricas.

f.2) ***Posibles utilidades en la instrucción de las causas***

f.2.a) Aumento de la eficiencia en el análisis de la escritura manuscrita y las firmas

En la actualidad, en ciertas ocasiones la prueba clave de un caso resulta ser un texto manuscrito o un documento rubricado, siendo fundamental determinar quién es su autor para averiguar así qué persona/s está/n tras la comisión del acto delictivo investigado.

En tales supuestos, pueden darse dos escenarios: por un lado, puede ocurrir que existan sospechas de que dicha escritura o dicha rúbrica pertenezcan a una persona en concreto, que a su vez puede o bien reconocerlo o bien negarlo; y, por otro lado, puede ocurrir que no exista indicio alguno de quién puede ser el autor de tal texto o firma.

En ambos casos, el uso de sistemas de IA de reconocimiento de escritura y firma sería de enorme utilidad.

Por un lado, respecto de los sistemas de IA de reconocimiento de escritura, procede poner de manifiesto que ya existen *software* capaces de analizar y comparar, a través de algoritmos entrenados para ello, imgánenes digitalizadas de escritura manuscrita, siendo que lo que hacen es hallar parámetros o patrones, detectar similitudes y coincidencias y aportar posibles candidatos, en caso de hallarlos, lo cual puede resultar tremendamente útil.

Si bien la policía española en la actualidad todavía no emplea tal clase de sistemas, que se hallan en fase de pruebas y experimentación, otros países, tales como Estados Unidos y Alemania, sí hacen ya uso de los mismos, especialmente para llevar a cabo investigaciones en materia terrorista. Entre otros, cabe destacar el sistema FISH ("*Forensic Information System for Handwriting*"), que contiene una enorme base de datos de imágenes de texto indubitadas digitalizadas con el fin de que puedan ser comparadas con otras, dubitadas, con fines de investigación criminal.[666]

Así, imaginemos que un político recibe en su oficina una carta manuscrita que contiene múltiples amenazas. Pensemos, asimismo, que tras ser analizada, no se hallan vestigios dactilares, siendo la escritura la única prueba que puede llevar al autor de los hechos. Supongamos, pues, que la policía decide digitalizar la carta e introducirla en un sistema de IA de reconocimiento de escritura para ser comparada con aquellas imágenes de textos manuscritos contenidas en su base de datos, aportando este un posible candidato (presunto autor de los hechos), que resultará investigado y será posteriormente sometido a un cuerpo de escritura por los expertos forenses para verificar o descartar su identidad.

Y es que en tal supuesto, de no haberse empleado un sistema de IA de reconocimiento automático de escritura, lo más probable hubiera sido que el caso hubiera quedado sobreseído por desconocimiento de autor y hubiera dado lugar a una ardua y compleja investigación policial, con el consumo de recursos que ello conlleva y con la sensación de desprotección que, con toda seguridad, se hubiera generado en la víctima.

[666] Véase National Institute of Standards and Technology, 2017.

Por otro lado, no obstante, respecto de los sistemas de IA de reconocimiento de firma, procede poner de manifiesto que su utilidad es mucho más limitada, al menos en la actualidad. Y es que, tal y como me explicó Guillermo Puerto, inspector jefe y responsable de la Sección de Documentoscopia del Cuerpo Nacional de Policía, la firma es una representación gráfica breve, que contiene muchísima menos información que la que alberga un cuerpo de escritura, por lo que no resultaría posible su digitalización y posterior análisis y comparación con otras, a diferencia de lo que ocurre con el reconocimiento de escritura.

Así, lo único que podría funcionar en tal ámbito (y, de hecho, muy bien y con mucha precisión, tal y como me aseguró el mencionado inspector) sería el análisis y la comparación de firmas plasmadas en soportes capaces de captar datos biométricos, habida cuenta de que estos detectan las múltiples características y parámetros que puede tener una rúbrica (forma, presión, velocidad, aceleración, etc) de forma mucho más amplia que los métodos de análisis forense tradicionales que analizan y comparan firmas estáticas. El problema, sin embargo, es que en la actualidad pocas veces los ciudadanos plasman sus rúbricas en tal tipo de soportes, si bien ello cada vez es más frecuente.

En cualquier caso, imaginemos que una señora presenta una denuncia en la policía por usurpación de identidad, habida cuenta de que ha hallado en casa de su hijo, con facultades mentales mermadas, un contrato de compraventa de un vehículo de lujo por sesenta mil euros en que consta una firma que, según dice, no ha sido realizada por él. Pensemos, además, que la mencionada madre de la víctima sospecha de su otro hijo, con características físicas muy similares al mismo, si bien este, al ser interrogado por la policía, niega tener relación alguna con los hechos. Figurémonos, asimismo, que el autor de los hechos acudió al concesionario de coches con la documentación del incapaz (todavía no incapacitado), haciéndose pasar por él, y plasmó la rúbrica para la financiación del vehículo en un soporte capaz de captar información biométrica. Ante tal circunstancia, supongamos que la policía decide someter a análisis la rúbrica del mencionado contrato mediante un sistema de IA de reconocimiento de firma, para lo cual, solicita a la víctima y a su hermano que plasmen sus rúbricas en un soporte también capaz de captar datos biométricos, de modo que las introduce en la base de datos y procede a compararlas con la contenida en el

contrato objeto de controversia, arrojando un resultado revelador: algunos de los rasgos de la firma falsificada coinciden con los de la firma del investigado.

En tal caso, si la firma falsificada únicamente hubiera constado en papel, el juez de instrucción habría tenido que acordar una prueba pericial caligráfica para que un experto forense procediera a su análisis y comparación, si bien los resultados hubieran resultado menos fiables puesto que no hubieran podido tenerse en cuenta tantos rasgos de la firma como en el caso de realizar su estudio comparando datos biométricos a través de la IA. Además, la práctica de una prueba pericial siempre resulta compleja y alarga la instrucción de las causas, por lo que la aplicación de un sistema de IA de reonocimiento de firma automático sin duda acortaría los plazos y revertiría en beneficio de la instrucción de la causa.

A lo ya expuesto, además, hay que añadir el papel disuasorio de este tipo de sistemas (puesto que, evidentemente, los delincuentes lo último que desean es ser descubiertos) que, asimismo, facilitan la detección precoz de los actos delictivos (tales como estafas, usurpaciones de identidad, etc) cometidos por aquellos que todavía no conozcan las bonanzas de los mismos y, asimismo, su ulterior investigación.

Y es que, imaginemos que un individuo acude a una entidad bancaria haciéndose pasar por otra persona y se dispone a firmar para contratar un préstamo en un soporte capaz de captar datos biométricos. Pensemos, asimismo, que dicho banco ya tenía registrada en su sistema la firma del cliente real, que en su momento también la plasmó en un soporte capaz de detectar datos biométricos, por lo que, en el mismo momento en que el usurpador procede a falsificar su firma, salta una alarma. Figurémonos, así, que el trabajador del banco, al percibir dicho aviso, procede a alertar a la policía y a intentar entretener al delincuente haciéndole consultas varias hasta que llegan los agentes a la sucursal bancaria y proceden a su detención.

En virtud de ello, en tal caso, con el uso de un sistema de IA de reconocimiento de firma, no solo se hubiera evitado la consumación de uno o varios delitos sino que además se habrían obtenido de forma inmediata pruebas bastantes para incriminar al autor de los hechos, sin necesidad de llevar a cabo más diligencias de investigación, quedando pues la instrucción de la causa reducida a la mínima expresión, con el ahorro de recursos materiales

y humanos que ello conlleva para la Administración de Justicia y todas las personas implicadas en la causa.

Y a ello hemos de sumar que, si bien un mero documento firmado de forma tradicional no contiene, en principio, ningún elemento que permita averiguar dónde y cuándo fue realmente plasmada la rúbrica (ya que es un acto de fe creer que esta tuvo lugar donde se indica a pie de página, por ejemplo), un documento firmado de forma digital o, en su caso, con firma digitalizada, contiene muchísima información al respecto.

Así, por un lado, en el archivo digital que conforma un documento firmado con un dispositivo de captación de firma, se guardan cifrados y encriptados multitud de datos, no solo de la biometría de la firma (huella biométrica) sino muchos otros referentes a las condiciones del acto de la extensión de la firma como son la fecha, la hora y la zona geográfica (información de la operación), en función de la IP donde se realizó la conexión del dispositivo. Y no solo eso, ya que tales datos están protegidos por una clave pública y otra privada que custodia una autoridad de servicios de certificación, garantizando con ello la integridad y la confidencialidad del documento, así como el no repudio y la no reutilización de la firma insertada en ese documento en dicha fecha.

Por otro lado, un archivo digital (como un "pdf", por ejemplo) firmado de forma electrónica[667] con los certificados contenidos en una tarjeta personal (entre otros, el DNI electrónico alberga asimismo cifrados los datos del acto de firma (lugar y hora, dispositivo utilizado y la IP desde donde se conecta el mismo y se produce la comprobación de los certificados contenidos en la tarjeta).

A.3. *Regulación europea y española*

Una vez analizadas las múltiples utilidades y bonanzas que los distintos sistemas de reconocimiento biométrico podrían aportar a la instrucción de las causas penales, procede hacer referencia a la otra cara de la moneda: la de los riesgos jurídicos que estos pueden entrañar.

[667] Que, tal y como prevé la Ley 6/2020, de 11 de noviembre, reguladora de determinados aspectos de los servicios electrónicos de confianza, tiene la misma validez y efectos que la firma manuscrita tradicional.

Para ello, no obstante, y con el fin de sistematizar la exposición, entiendo oportuno, en primer lugar, hacer referencia a la actual regulación europea y española aplicable a los mencionados sistemas de IA, con el fin de dar a conocer cuál es la base de la que partimos y qué instrumentos legales deberían tener en cuenta tanto el legislador como, en su caso, los operadores jurídicos, para legitimar o determinar la legalidad (o no) de las herramientas analizadas; y, en segundo lugar, considero necesario hacer alusión a los potenciales riesgos jurídicos que, en general, puede implicar el uso de tales tecnologías.

En relación con ello es importante tener claro que el instrumento legal específico que a día de hoy procede tener en cuenta para determinar la viabilidad jurídica o, en su caso, legitimar el uso de los sistemas de IA que emplean datos biométricos por parte de las autoridades competentes para fines de investigación criminal es la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales.[668]

De forma supletoria, no obstante, en virtud de lo dispuesto en el artículo 6.2 de la mencionada LO 7/2021, de 26 de mayo, cuando los datos personales recogidos por las antedichas autoridades competentes sean tratados para otros fines distintos de los establecidos en el artículo 1 de tal cuerpo legal[669] (y siempre en caso de que dicho tratamiento esté autorizado por el Derecho de la Unión Europea o por la legislación española), serán de aplicación el Reglamento General de Protección de Datos (RGPD) y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, a menos que el tratamiento se efectúe como parte de una actividad que quede fuera del ámbito de aplicación del Derecho de la Unión Europea.

[668] Que, tal y como ya se ha expuesto con anterioridad, ha venido a transponer la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos.

[669] A saber, "*fines de prevención, detección, investigación y enjuiciamiento de infracciones penales o de ejecución de sanciones penales, incluidas la protección y prevención frente a las amenazas contra la seguridad pública."*

No obstante, evidentemente, y de forma general, el uso de los mencionados sistemas de IA está sujeto asimismo a lo dispuesto en la Carta de Derechos Fundamentales de la UE, en la Constitución Española de 1978 y en el resto de legislación aplicable a cada caso concreto.

Tal y como se desprende de lo dispuesto en el artículo 2.1 de la mencionada LO 7/21, de 26 de mayo, dicho cuerpo legal será de aplicación "*al tratamiento total o parcialmente automatizado de datos personales, así como al tratamiento no automatizado de datos personales contenidos o destinados a ser incluidos en un fichero, realizado por las autoridades competentes, con fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, incluidas la protección y prevención frente a las amenazas contra la seguridad pública*." Y, en relación con ello, dispone el apartado 2 de tal precepto que: "*El tratamiento de los datos personales llevado a cabo con ocasión de la tramitación por los órganos judiciales y fiscalías de las actuaciones o procesos de los que sean competentes, así como el realizado dentro de la gestión de la Oficina judicial y fiscal, en el ámbito del artículo 1, se regirá por lo dispuesto en la presente Ley Orgánica, sin perjuicio de las disposiciones de la Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial, las leyes procesales que le sean aplicables y, en su caso, por la Ley 50/1981, de 30 de diciembre, por la que se regula el Estatuto Orgánico del Ministerio Fiscal. Las autoridades de protección de datos a las que se refiere el capítulo VI* (a saber, las Autoridades de Protección de Datos Independientes) *no serán competentes para controlar estas operaciones de tratamiento*.", quedando fuera del ámbito de aplicación de la antedicha ley los tratamientos de datos personales previstos en el apartado 3, a saber:

"*a) Los realizados por las autoridades competentes para fines distintos de los previstos en el artículo 1, incluidos los fines de archivo por razones de interés público, investigación científica e histórica o estadísticos. Estos tratamientos se someterán plenamente a lo establecido en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos), así como en la Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales.*

b) Los llevados a cabo por los órganos de la Administración General del Estado en el marco de las actividades comprendidas en el ámbito de aplicación del capítulo II del título V del Tratado de la Unión Europea.

c) Los tratamientos que afecten a actividades no comprendidas en el ámbito de aplicación del Derecho de la Unión Europea.

d) Los sometidos a la normativa sobre materias clasificadas, entre los que se encuentran los tratamientos relativos a la Defensa Nacional.

e) Los tratamientos realizados en las acciones civiles y procedimientos administrativos o de cualquier índole vinculados con los procesos penales que no tengan como objetivo directo ninguno de los fines del artículo 1."

Sentado lo anterior, en primer lugar, como cuestión previa, resulta necesario poner de manifiesto que tanto la legislación europea como la legislación española de datos otorgan a los datos biométricos la calificación de datos personales, por resultar especialmente sensibles y, por ende, los sujeta a un elevado grado de protección. Ello, bajo mi punto de vista, tiene su principal fundamento, tal y como se deduce de lo dispuesto por la Agencia Española de Protección de Datos (que en junio de 2020 publicó un documento aclarando catorce puntos relacionados con el uso de los sistemas de reconocimiento facial), en el hecho de que "*A diferencia de una contraseña o un certificado, los datos biométricos recogidos durante un procedimiento de autenticación o identificación revelan más información personal sobre el sujeto. Dependiendo de los datos biométricos recogidos, pueden derivarse datos del sujeto como su raza o género, su estado emocional, enfermedades, discapacidades y características genéticas, consumos de sustancias, etc*".[670]

[670] AEPD, 2020.

En virtud de tal calificación, por un lado, resultan de aplicación respecto de los datos biométricos, los principios relativos al tratamiento de datos personales enumerados en el artículo 6.1 de la LO 7/21, de 26 de mayo, que dispone que:

"1. *Los datos personales serán:*
a) Tratados de manera lícita y leal.
b) Recogidos con fines determinados, explícitos y legítimos, y no serán tratados de forma incompatible con esos fines.
c) Adecuados, pertinentes y no excesivos en relación con los fines para los que son tratados.
d) Exactos y, si fuera necesario, actualizados. Se adoptarán todas las medidas razonables para que se supriman o rectifiquen, sin dilación indebida, los datos personales que sean inexactos con respecto a los fines para los que son tratados.
e) Conservados de forma que permitan identificar al interesado durante un período no superior al necesario para los fines para los que son tratados.
f*) Tratados de manera que se garantice una seguridad adecuada, incluida la protección contra el tratamiento no autorizado o ilícito y contra su pérdida, destrucción o daño accidental. Para ello, se utilizarán las medidas técnicas u organizativas adecuadas*."

Asimismo, en su apartado 3 se permite que los datos personales puedan ser tratados por el mismo responsable o por otro, para alguno de los fines previstos en el artículo 1 distintos de aquel para el que hubieran sido recogidos, en la medida en que concurran cumulativamente las dos circunstancias siguientes:

"*a) Que el responsable del tratamiento sea competente para tratar los datos para ese otro fin, de acuerdo con el Derecho de la Unión Europea o la legislación española.*

b) Que el tratamiento sea necesario y proporcionado para la consecución de ese otro fin, de acuerdo con el Derecho de la Unión Europea o la legislación española."[671]

[671] Asimismo, tal y como establece el apartado 4, "*El tratamiento por el mismo responsable o por otro podrá incluir el archivo por razones de interés público, y el uso científico, estadístico o histórico para los fines establecidos en el artículo 1, con sujeción a las garantías adecuadas para los derechos y libertades de los interesados*."

Y, con el propósito de garantizar el cumplimiento de los requisitos legalmente establecidos, el apartado 5 del mencionado precepto establece una obligación para el responsable del tratamiento de datos: deberá garantizar y estar en condiciones de demostrar el cumplimiento de los requisitos establecidos en el mismo.

Por otro lado, resulta de aplicación la prohibición general prevista en el artículo 11 de la mencionada LO 7/21, de 26 de mayo, que establece que: "*1. El tratamiento sólo será lícito en la medida en que sea necesario para los fines señalados en el artículo 1 y se realice por una autoridad competente en ejercicio de sus funciones.*

2. Cualquier ley que regule tratamientos de datos personales para los fines incluidos dentro del ámbito de aplicación de esta Ley Orgánica deberá indicar, al menos, los objetivos del tratamiento, los datos personales que vayan a ser objeto del mismo y las finalidades del tratamiento."

Asimismo, es aplicable el deber de colaboración establecido en el artículo 7, que dispone:

"*1. Las Administraciones públicas, así como cualquier persona física o jurídica, proporcionarán a las autoridades judiciales, al Ministerio Fiscal o a la Policía Judicial los datos, informes, antecedentes y justificantes que les soliciten y que sean necesarios para la investigación y enjuiciamiento de infracciones penales o para la ejecución de las penas. La petición de la Policía Judicial se deberá ajustar exclusivamente al ejercicio de las funciones que le encomienda el artículo 549.1 de la Ley Orgánica 6/1985, de 1 de julio y deberá efectuarse siempre de forma motivada, concreta y específica, dando cuenta en todo caso a la autoridad judicial y fiscal.*

La comunicación de datos, informes, antecedentes y justificantes por la Administración Tributaria, la Administración de la Seguridad Social y la Inspección de Trabajo y Seguridad Social, se efectuará de acuerdo con su legislación respectiva.

2. En los restantes casos, las Administraciones públicas, así como cualquier persona física o jurídica, proporcionarán los datos, informes, antecedentes y justificantes a las

autoridades competentes que los soliciten, siempre que estos sean necesarios para el desarrollo específico de sus misiones para la prevención, detección e investigación de infracciones penales y para la prevención y protección frente a un peligro real y grave para la seguridad pública. La petición de la autoridad competente deberá ser concreta y específica y contener la motivación que acredite su relación con los indicados supuestos.

3. No será de aplicación lo dispuesto en los apartados anteriores cuando legalmente sea exigible la autorización judicial para recabar los datos necesarios para el cumplimiento de los fines del artículo 1.

4. En los supuestos contemplados en los apartados anteriores, el interesado no será informado de la transmisión de sus datos a las autoridades competentes, ni de haber facilitado el acceso a los mismos por dichas autoridades de cualquier otra forma, a fin de garantizar la actividad investigadora.

Con el mismo propósito, los sujetos a los que el ordenamiento jurídico imponga un deber específico de colaboración con las autoridades competentes para el cumplimiento de los fines establecidos en el artículo 1, no informarán al interesado de la transmisión de sus datos a dichas autoridades, ni de haber facilitado el acceso a los mismos por dichas autoridades de cualquier otra forma, en cumplimiento de sus obligaciones específicas."

En adición, resultan aplicables los plazos de conservación y revisión previstos en el artículo 8 que dispone como regla general que el responsable del tratamiento de los datos personales determinará que la conservación de estos tenga lugar sólo durante el tiempo necesario para cumplir con los fines de prevención, detección, investigación y enjuiciamiento de infracciones penales o de ejecución de sanciones penales, incluidas la protección y prevención frente a las amenazas contra la seguridad pública, estableciendo un plazo máximo para la supresión de los datos de veinte años, salvo que concurran factores como la existencia de investigaciones abiertas o delitos que no hayan prescrito, la no conclusión de la ejecución de la pena, la reincidencia, la necesidad de protección de las víctimas u otras circunstancias motivadas que hagan necesario el tratamiento de los datos para el cumplimiento de los antedichos fines. Asimismo, el mencionado precepto impone una obligación al responsable del tratamiento de los datos personales: la revisión de la

necesidad de conservar, limitar o suprimir el conjunto de estos contenidos en cada una de las actividades de tratamiento bajo su responsabilidad, como máximo cada tres años, atendiendo especialmente en cada revisión a la edad del afectado, el carácter de los datos y a la conclusión de una investigación o procedimiento penal (y si es posible, se hará mediante el tratamiento automatizado apropiado).

Y a esta hay que sumar otras dos obligaciones que la ley (artículos 9 y 10, respectivanente) imponen al referido responsable del tratamiento de datos personales, a saber:

-establecimiento de distinciones, en la medida de lo posible, entre los datos personales de las distintas categorías de interesados (sin que ello deba impedir la aplicación del derecho a la presunción de inocencia tal como lo garantiza el artículo 24 de la Constitución), tales como:

"*a) Personas respecto de las cuales existan motivos fundados para presumir que hayan cometido, puedan cometer o colaborar en la comisión de una infracción penal.*
b) Personas condenadas o sancionadas por una infracción penal.
c) Víctimas o afectados por una infracción penal o que puedan serlo.
d) Terceros involucrados en una infracción penal como son: personas que puedan ser citadas a testificar en investigaciones relacionadas con infracciones o procesos penales ulteriores, personas que puedan facilitar información sobre dichas infracciones, o personas de contacto o asociados de una de las personas mencionadas en las letras a) y b)."

-establecimiento, en la medida de lo posible, de una distinción entre los datos personales basados en hechos y los basados en apreciaciones personales.

Y el artículo 10, además, en su apartado 2 dispone una obligación de verificación de calidad de los datos: "*Las autoridades competentes adoptarán todas las medidas razonables para garantizar que los datos personales que sean inexactos, incompletos o no estén actualizados, no se transmitan ni se pongan a disposición de terceros. En toda transmisión de datos se trasladará al mismo tiempo la valoración de su calidad, exactitud y actualización.*

En la medida de lo posible, en todas las transmisiones de datos personales se añadirá la información necesaria para que la autoridad competente receptora pueda valorar hasta qué punto son exactos, completos y fiables, y en qué medida están actualizados. Igualmente, la autoridad competente transmisora, en la medida en que sea factible, controlará la calidad de los datos personales antes de transmitirlos o ponerlos a disposición de terceros.

3. Si se observara que los datos personales transmitidos son incorrectos o que se han transmitido ilegalmente, estas circunstancias se pondrán en conocimiento del destinatario sin dilación indebida. En tal caso, los datos deberán rectificarse o suprimirse, o el tratamiento deberá limitarse de conformidad con lo previsto en el artículo 23."

De forma específica, no obstante, es importante poner el foco en lo dispuesto en el artículo 13 respecto del tratamiento de categorías especiales de datos personales (entre los que se hallan los datos biométricos dirigidos a identificar de manera unívoca a una persona física), al que ya se hizo especial mención al hablar de las herramientas de IA de evaluación de riesgos[672]. Y es que en este se prevé una prohibición general de tratamiento de tal clase de datos, con tales fines, salvo en determinadas excepciones previstas como *numerus clausus* en el mencionado precepto. Así, el citado artículo 13 dispone (con subrayado propio):

"*El tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas o la afiliación sindical, así como el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, los datos relativos a la salud o a la vida sexual o a la orientación sexual de una persona física, sólo se permitirá cuando sea estrictamente necesario, con sujeción a las garantías adecuadas para los derechos y libertades del interesado y cuando se cumplan alguna de las siguientes circunstancias:*

a) Se encuentre previsto por una norma con rango de ley o por el Derecho de la Unión Europea.

[672] Véase pág. 213.

b) Resulte necesario para proteger los intereses vitales, así como los derechos y libertades fundamentales del interesado o de otra persona física.
c) Dicho tratamiento se refiera a datos que el interesado haya hecho manifiestamente públicos.

2. Las autoridades competentes, en el marco de sus respectivas funciones y competencias, podrán tratar datos biométricos dirigidos a identificar de manera unívoca a una persona física con los fines de prevención, investigación, detección de infracciones penales, incluidas la protección y la prevención frente a las amenazas contra la seguridad pública.

3. Los datos de los menores de edad y de las personas con capacidad modificada judicialmente o que estén incursas en procesos de dicha naturaleza, se tratarán garantizando el interés superior de los mismos y con el nivel de seguridad adecuado."

De acuerdo con lo expuesto, las autoridades únicamente podrán utilizar sistemas de IA que empleen datos biométricos para identificar de manera unívoca a una persona física cuando ello sea estrictamente necesario, con observancia de las garantías adecuadas para los derechos y libertades del interesado y cuando ello se halle previsto por una norma con rango de ley o por el Derecho de la UE (lo cual hay que ir verificando en cada momento y en cada caso, en relación con cada una de las clases de datos biométricos, debiendo prestar especial atención a lo que se disponga en el texto definitivo del Reglamento sobre IA que está en ciernes); cuando ello resulte necesario para proteger los intereses vitales, así como los derechos y libertades fundamentales del interesado o de otra persona física (por ejemplo, en caso de desaparición de un menor o de un individuo adulto); cuando dicho tratamiento se refiera a datos que el interesado haya hecho manifiestamente públicos, lo cual se entiende como un consentimiento para su tratamiento; o con fines de prevención, investigación y detección de infracciones penales, incluidas la protección y la prevención frente a las amenazas contra la seguridad pública (en el caso de una amenaza terrorista concreta, por ejemplo).

Y es que respecto de esto último es importante puntualizar que no todo tratamiento de datos biométricos destinado a identificar de modo unívoco a una persona con fines de prevención, investigación, detección de infracciones penales, puede quedar legitimado bajo

el paraguas de la seguridad pública, habida cuenta de que el precepto debe interpretarse de forma restrictiva (como todos aquellos que hacen referencia a la limitación de derechos fundamentales) y además debe cumplir con el resto de la legislación vigente aplicable, que prohíbe las investigaciones prospectivas.

Así, por ejemplo, si la policía recibe un aviso de que hay un terrorista preparado para aparcar un vehículo cargado de explosivos en una céntrica calle comercial, o de que hay una persona con un arma de fuego dentro de un campo de fútbol que se dirige a disparar a los hinchas del equipo contrario, desde luego entiendo que la utilización por parte de las autoridades de un sistema de reconocimiento facial en tiempo real, por ejemplo, quedaría plenamente legitimada, al tratarse de una amenaza concreta para la seguridad pública. No obstante, considero que, de modo genérico, y a pesar de la amenaza terrorista constante que pesa sobre la UE o del evidente riesgo existente en los partidos de fútbol de mayor rivalidad, no resultaría ni proporcionado ni estrictamente necesario y, por ende, devendría contrario a los derechos y libertades de los ciudadanos, la utilización preventiva de tal clase de herramientas con el fin de localizar posibles personas clasificadas como sospechosas, registradas como violentas, etc (lo cual, sin embargo, sí resulta legítimo en China, donde los estándares de protección de los derechos y libertades individualles son significativamente menores que en la UE) deambulando por la vía pública o por un campo de fútbol, a los efectos de detectar la posible comisión de un ilícito penal.

Y es que sin duda hay que llevar a cabo una valoración de las circunstancias en cada caso concreto en que se pretenda, por parte de las autoridades, llevar a cabo un tratamiento de datos biométricos con el fin de identificar de modo unívoco a una persona física, a los efectos de determinar si este podría ser subsumido o no en lo dispuesto en el mencionado artículo 13, para evitar así vulneraciones de derechos. Dicho esto, y siendo que en dicho precepto se contienen ciertos conceptos jurídicos indeterminados (a saber, “estrictamente necesario”, “manifiestamente públicos”, “seguridad pública”, etc), lo cierto es que considero que pueden darse discrepancias jurídicas lógicas en relación con lo expuesto, por lo que los tribunales son los que, en ultima instancia, tendrán la palabra.

No obstante, todo ello debe ser puesto en relación, además, con lo previsto en el artículo 14 de la citada ley, que tal y como ya se expuso al hacer referencia a las herramientas de

IA de evaluación de riesgos[673], en su apartado 1 prohíbe terminantemente, de forma concreta, aquellas decisiones basadas únicamente en un tratamiento automatizado de datos personales, incluida la elaboración de perfiles, que produzcan efectos jurídicos negativos para el interesado o que le afecten significativamente, salvo que se autorice expresamente por una norma con rango de ley o por el Derecho de la Unión Europea (que deberá establecer las medidas adecuadas para salvaguardar los derechos y libertades del interesado, incluyendo el derecho a obtener la intervención humana en el proceso de revisión de la decisión adoptada), lo cual entiendo que es perfectamente prudente y absolutamente necesario para salvaguardar los derechos y libertades de los ciudadanos y, sobre todo, limitar la autonomía de las máquinas.

Ello, sin embargo, como el propio precepto indica, hace solamente referencia a aquellas decisiones basadas únicamente en un tratamiento automatizado de datos personales, incluida la elaboración de perfiles, por lo que no resultaría de aplicación respecto de aquellas decisiones tomadas por las autoridades con base en el resultado de un tratamiento automatizado de datos personales, incluida la elaboración de perfiles, en caso de que además hubiera habido una intervención humana cualificada. Me explico.

Bajo mi punto de vista, no es lo mismo el hecho de que una cámara con un sistema de reconocimiento facial, por ejemplo, capte la presencia ilegítima en un determinado territorio de un individuo con prohibición de entrada y de forma automática se ordene la remisión de una citación como investigado por un presunto delito de quebrantamiento de condena sin más control humano, que el hecho de que el mencionado sistema, en caso de detectar a tal sujeto, haga saltar una alerta dirigida a la policía y esta mande una patrulla al lugar de los hechos para efectuar comprobaciones, hable con posibles testigos, acuda a su domicilio en caso de no hallarlo, etc y finalmente concluya que existen indicios para remitir un atestado al juzgado con todos los indicios disponibles (incluido el resultado arrojado por el sistema de reconocimiento facial) y, desde este, se decida citar al individuo como investigado por la presunta comisión de un delito de quebrantamiento de condena.

[673] Véanse págs. 210-213.

Y es que en mi opinión, las decisiones policiales o judiciales tomadas con base en los resultados arrojados por los sistemas de IA de investigación criminal que emplean datos biométricos con el fin de identificar de modo unívoco a una persona, siempre y cuando haya además una intervención humana cualificada (como ocurre hoy en día prácticamente en el 100% de los casos), deben ser sin duda calificadas como decisiones legítimas, aunque produzcan efectos jurídicos negativos para el interesado o le afecten significativamente (puesto que de ello depende su implicación o no en el asunto investigado, por ejemplo).

Además, respecto de las decisiones analizadas, aplica también lo dispuesto en el apartado 2 del mencionado artículo 14 que dispone que "*las decisiones a las que se refiere el apartado anterior no se basarán en las categorías especiales de datos personales contempladas en el artículo 13* (a saber, datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas o la afiliación sindical, así como el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, los datos relativos a la salud o a la vida sexual o a la orientación sexual de una persona física), *salvo que se hayan tomado las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.*", lo cual deberá ser tenido en cuenta por el legislador a la hora de dictar la norma que permita su uso, quedando prohibida en cualquier caso la elaboración de perfiles que dé lugar a una discriminación de las personas físicas sobre la base de las mencionadas categorías especiales de datos personales.

En la actualidad, no obstante, salvo error u omisión, no hay vigente ninguna ley (ni española ni europea) que autorice de forma expresa las decisiones basadas únicamente en un tratamiento automatizado de datos personales, incluida la elaboración de perfiles, que produzcan efectos jurídicos negativos para el interesado o que le afecten significativamente, para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales o de ejecución de sanciones penales, por lo que bajo mi punto de vista, al menos por el momento, el uso de esta clase de herramientas por parte de las autoridades en los términos expuestos no resultaría legítimo.

Y es que no entiendo que para estos supuestos pudiera ser de aplicación el RGPD, habida cuenta de que, si bien la Disposición Adicional Primera de la LO 7/21, de 26 de mayo,

dispone (con subrayado propio) que "*1. El tratamiento de los datos personales procedentes de las imágenes y sonidos obtenidos mediante la utilización de cámaras y videocámaras por las Fuerzas y Cuerpos de Seguridad, por los órganos competentes para la vigilancia y control en los centros penitenciarios y para el control, regulación, vigilancia y disciplina del tráfico, para los fines previstos en al artículo 1, se regirá por esta Ley Orgánica, sin perjuicio de los requisitos establecidos en regímenes legales especiales que regulan otros ámbitos concretos como el procesal penal, la regulación del tráfico o la protección de instalaciones propias. 2. Fuera de estos supuestos, dichos tratamientos se regirán por su legislación específica y supletoriamente por el Reglamento (UE) 2016/679 y por la Ley Orgánica 3/2018, de 5 de diciembre*.", lo cierto es que falta regulación específica y el RGPD, por un lado, dispone especialmente en su artículo 2.2.d) que no aplica al tratamiento de datos personales efectuado por parte de las autoridades competentes con fines de prevención, investigación, detección o enjuiciamiento de infracciones penales, o de ejecución de sanciones penales, incluida la de protección frente a amenazas a la seguridad pública y su prevención (lo cual, no obstante, resulta contradictorio); y, por otro lado, no prevé de forma expresa el tratamiento total o parcialmente automatizado de datos personales y la viabilidad jurídica de las decisiones tomadas con base en ello, por lo que no puede considerarse norma habilitadora en los términos previstos en el artículo 14 de la LO 7/21, de 26 de mayo.

Ello sin perjuicio, por supuesto, de lo que pueda ocurrir en caso de que se dicte una norma nacional o europea que así lo autorice en la forma prevista en el artículo 14 de la mencionada LO 7/21, de 26 de mayo, lo cual entiendo que va a tener lugar en un futuro próximo, con toda probabilidad, cuando entre en vigor el Reglamento de IA que está siendo elaborado por las instituciones europeas, ya que actualmente solo se cuenta con la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión, que tal y como ya se ha dicho en ocasiones anteriores, por ahora es una mera propuesta. Y, en relación con ello, entiendo que de acuerdo con lo dispuesto en el artículo 6.2 de la mencionada Propuesta de la Comisión, ciertas de las herramientas de IA analizadas que emplean datos biométricos

tendrán además la consideración de sistemas de IA de alto riesgo, por lo que les resultará de aplicación lo dispuesto para estos de forma específica.[674]

Así, por un lado, aquellas herramientas de IA que empleen datos biométricos y no sean consideradas de alto riesgo, deberán cumplir con lo dispuesto en el Título IV de la mencionada Propuesta, que establece una serie de "*Obligaciones de transparencia para determinados sistemas de IA*" y dispone:

"*1. Los proveedores garantizarán que los sistemas de IA destinados a interactuar con personas físicas estén diseñados y desarrollados de forma que dichas personas estén informadas de que están interactuando con un sistema de IA, excepto en las situaciones en las que esto resulte evidente debido a las circunstancias y al contexto de utilización. Esta obligación no se aplicará a los sistemas de IA autorizados por la ley para fines de detección, prevención, investigación o enjuiciamiento de infracciones penales, salvo que estos sistemas estén a disposición del público para denunciar una infracción penal.*

2. Los usuarios de un sistema de reconocimiento de emociones o de un sistema de categorización biométrica informarán del funcionamiento del sistema a las personas físicas expuestas a él. Esta obligación no se aplicará a los sistemas de IA utilizados para la categorización biométrica autorizados por la ley para fines de detección, prevención e investigación de infracciones penales. Los usuarios de un sistema de IA que genere o manipule contenido de imagen, sonido o vídeo que se asemeje notablemente a personas, objetos, lugares u otras entidades o sucesos existentes, y que pueda inducir erróneamente a una persona a pensar que son auténticos o verídicos (ultrafalsificación), harán público que el contenido ha sido generado de forma artificial o manipulado.

No obstante, el primer párrafo no se aplicará cuando el uso esté legalmente autorizado por la ley para fines de detección, prevención, investigación y enjuiciamiento de infracciones penales.

[674] Véanse págs.192-197.

4. Los apartados 1, 2 y 3 no afectarán a los requisitos y obligaciones dispuestos en el título III del presente Reglamento."

Y es que, de acuerdo con lo expuesto, lo que se fija principalmente es una obligación general de información a los interesados que interactúen o se vean afectados por el uso de un sistema de IA que, en este caso, emplee datos biométricos, pero se establece una excepción para ello: que la utilización de tales herramientas esté legalmente autorizada por la ley para fines de detección, prevención, investigación y enjuiciamiento de infracciones penales, lo cual entiendo que es absolutamente necesario, adecuado y proporcionado, puesto que en caso contrario ello podría frustrar justamente los objetivos legítimamente perseguidos por las autoridades en tales casos.

Y, por otro lado, aquellas herramientas de IA que empleen datos biométricos y sean consideradas de alto riesgo, deberán regirse por lo dispuesto en el Título III de la mencionada Propuesta. En relación con ello, el ya citado artículo 6.2, que nombra los sistemas de alto riesgo, se remite a lo dispuesto en el Anexo III, que califica como tales las siguientes herramientas de interés analizadas en la presente Sección:

-en relación con la identificación biométrica y categorización de personas físicas:

"*a) los sistemas de IA destinados a utilizarse en la identificación biométrica remota «en tiempo real» o «en diferido» de personas físicas*"

-en relación con asuntos relacionados con la aplicación de la ley: (…)

"*b) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley como polígrafos y herramientas similares, o para detectar el estado emocional de una persona física*; (…)

f) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para la elaboración de perfiles de personas físicas, de conformidad con lo dispuesto en el artículo 3, apartado 4, de la Directiva (UE) 2016/680, durante la detección, la investigación o el enjuiciamiento de infracciones penales;

g) sistemas de IA destinados a utilizarse para llevar a cabo análisis sobre infracciones penales en relación con personas físicas que permitan a las autoridades encargadas de la aplicación de la ley examinar grandes conjuntos de datos complejos vinculados y no vinculados, disponibles en diferentes fuentes o formatos, para detectar modelos desconocidos o descubrir relaciones ocultas en los datos."

Respecto de ello, además, procede hacer especial mención a lo dispuesto en la mencionada Propuesta sobre los sistemas de identificación biométrica remota en tiempo real (considerados de alto riesgo, tal y como se acaba de exponer) que se definen en el artículo 3 apartados (36) como sistemas de IA "*con el propósito de identificar a personas físicas a distancia mediante la comparación de los datos biométricos de una persona con los datos biométricos contenidos en una base de datos de referencia, sin que el usuario del sistema conozca previamente si la persona estará presente y podrá ser identificada*", y (37), que concreta el significado de "en tiempo real", disponiendo que ello implica que "*la captación, la comparación y la identificación de datos biométricos ocurren sin un retraso significativo.*"

En relación con tal clase de sistemas, por un lado, con carácter general se prohíbe su uso en espacios públicos y con fines policiales. No obstante, se establecen una serie de excepciones (por ejemplo, para prevenir amenazas terroristas inminentes y concretas, para localizar y/o identificar a los sospechosos o a los autores de delitos graves, o buscar a un menor desaparecido), siendo necesaria en cualquier caso para su uso una autorización judicial (o de un organismo independiente) que haga una ponderación y establezca límites tanto temporales como geográficos, y relativos a las bases de datos que pueden ser empleadas.

Y es que el artículo 5 del mencionado texto legal dispone: "*Quedan prohibidas las siguientes prácticas de Inteligencia Artificial:* (…)

d) el uso de sistemas de identificación biométrica remota "en tiempo real" en espacios de acceso público con fines policiales, a menos que dicho uso sea estrictamente necesario para uno de los siguientes objetivos:

(i) la búsqueda selectiva de específicas posibles víctimas de delitos, incluyendo niños desaparecidos;

(ii) la prevención de una amenaza específica, sustancial e inminente para la vida o seguridad física de las personas físicas o de un ataque terrorista;

(iii) la detección, localización, identificación o enjuiciamiento del autor o sospechoso de la comisión de uno de los delitos mencionados en el Artículo 2 (2) de la Decisión Marco 2002/584/JHA, siempre que sea punible en el Miembro Estado afectado por una orden de detención o una pena privativa de libertad de al menos tres años de duración, según la ley de dicho Estado miembro."

En relación con ello, los delitos previstos en el artículo 2(2) de la mencionada Decisión Marco son los siguientes:

-pertenencia a organización delictiva,

-terrorismo,

-trata de seres humanos,

-explotación sexual de los niños y pornografía infantil,

-tráfico ilícito de estupefacientes y sustancias psicotrópicas,

-tráfico ilícito de armas, municiones y explosivos,

-corrupción,

-fraude, incluido el que afecte a los intereses financieros de las Comunidades Europeas con arreglo al Convenio de 26 de julio de 1995 relativo a la protección de los intereses financieros de las Comunidades Europeas,

-blanqueo del producto del delito,

-falsificación de moneda, incluida la falsificación del euro,

-delitos de alta tecnología, en particular delito informático,

-delitos contra el medio ambiente, incluido el tráfico ilícito de especies animales protegidas y de especies y variedades vegetales protegidas,

-ayuda a la entrada y residencia en situación ilegal,

-homicidio voluntario, agresión con lesiones graves,

-tráfico ilícito de órganos y tejidos humanos,

-secuestro, detención ilegal y toma de rehenes,

-racismo y xenofobia,

-robos organizados o a mano armada,

-tráfico ilícito de bienes culturales, incluidas las antigüedades y las obras de arte,

-estafa,

-chantaje y extorsión de fondos,

-violación de derechos de propiedad industrial y falsificación de mercancías,

-falsificación de documentos administrativos y tráfico de documentos falsos,

-falsificación de medios de pago,

-tráfico ilícito de sustancias hormonales y otros factores de crecimiento,

-tráfico ilícito de materiales radiactivos o sustancias nucleares,

-tráfico de vehículos robados,

-violación,

-incendio voluntario,

-delitos incluidos en la jurisdicción de la Corte Penal Internacional,

-secuestro de aeronaves y buques,

-sabotaje.

En el caso de España, procede poner de manifiesto, no obstante, que no todos los antedichos delitos cumplen con los requisitos penológicos establecidos en el transcrito artículo 5.d) y, por ende, en el momento en que entre en vigor el proyectado Reglamento de IA, salvo que haya cambios, los sistemas de identificación biométrica remota en tiempo real en espacios de acceso público no podrán ser empleados en nuestro país para la detección, localización, identificación o enjuiciamiento de los autores o sospechosos ni, por un lado, de la comisión de un delito de ayuda a la entrada y residencia en situación ilegal, previsto en el artículo 318bis del Código Penal, penado con pena de multa de tres a doce meses o prisión de tres

meses a un año; ni, por otro lado, de la comisión de un delito contra la propiedad industrial (tipo básico) previsto en el artículo 273 del Código Penal, penado con con pena de prisión de seis meses a dos años y multa de doce a veinticuatro meses.

Asimismo, el mencionado artículo 5, en su apartado 3, tal y como se ha avanzado, requiere la concurrencia de autorización judicial (o, en su caso, de una autoridad administrativa nacional independiente) razonada, salvo en casos de urgencia justificada (en que se permite una ratificación judicial o administrativa posterior), para llevar a cabo el uso de tales sistemas de identificación biométrica remota en tiempo real en espacios de acceso público en los casos previstos en el apartado 1.d) y, para ello, otorga una guía a los jueces y magistrados (o, en su caso, autoridades administrativas nacionales independientes), sobre qué circunstancias y elementos deben tener en cuenta para valorar en cada caso concreto la concesión o la denegación de la utilización de tales herramientas tecnológicas.

Así, el apartado 3, párrafo 2, dispone: "*La autoridad judicial o administrativa competente solo otorgará la autorización cuando, sobre la base de las pruebas objetivas o los indicios claros que se le presenten, el uso del sistema de identificación biométrica remota en "tiempo real" en cuestión resulte necesario y proporcionado para lograr uno de los objetivos especificados en apartado 1, letra d), que debe ser indicado en la solicitud. Al decidir sobre tal solicitud, la autoridad judicial o administrativa competente tendrá en cuenta los elementos mencionado en el apartado 2*."

Y es que, el apartado 2 de dicho artículo 5 dispone que deberán tenerse en cuenta:

"*a) la naturaleza de la situación que da lugar al posible uso, en particular la gravedad, la probabilidad y la entidad del daño que se causaría en caso de no emplear el sistema;*

(b) las consecuencias del uso del sistema para los derechos y libertades de todas las personas interesadas, y en particular la gravedad, la probabilidad y la entidad de tales consecuencias."

Y se añade: "*Además, el uso de sistemas de identificación biométrica remota en "tiempo real" en espacios de acceso público con el propósito de hacer cumplir la ley, para cualquiera de los objetivos a que se refiere el apartado 1, letra d), deberá cumplir con garantías y requisitos de necesariedad y proporcionalidad, en relación con dicho uso, en particular en lo que respecta a las limitaciones temporales, geográficas y personales*."

No obstante, el mencionado artículo 5, en su apartado 4, abre la puerta a que los Estados Miembros autoricen y regulen de forma específica el uso policial de tales sistemas para los fines descritos en el apartado 1.d), siempre dentro de los límites y en las condiciones enumeradas en dicho apartado 1.d) así como en los apartados 2 y 3. Y, en tal sentido, dispone: "*Cada Estado Miembro establecerá en su legislación nacional las normas necesarias para la solicitud, la emisión, el ejercicio y la supervisión de las autorizaciones a las que se refiere el apartado 3. Dichas normas también especificarán para cuál de los objetivos enumerados en el apartado 1, letra d), incluido para cuál de los delitos a los que se hace referencia en inciso iii) del mismo, las autoridades competentes podrán ser autorizadas para utilizar dichos sistemas con el propósito de hacer cumplir la ley*."

Por otro lado, siendo que tales sistemas son considerados de alto riesgo, tal y como se dispone en el artículo 6.2, en relación con el Anexo III, en los casos en que su uso esté permitido, este se halla sujeto a estrictos requisitos y obligaciones.

En relación con ello, sin embargo, las voces críticas no se han hecho esperar. Así, por ejemplo, desde Amnistía Internacional se ha denunciado la insuficiencia de la regulación propuesta, habida cuenta de que, según expresa Rasha Abdul Rahim, directora de Amnesty Tech: "*La propuesta de la UE no está ni mucho menos a la altura de lo que se necesita para mitigar el enorme abuso potencial de tecnologías como los sistemas de reconocimiento facial. Según la prohibición propuesta, la policía seguirá estando facultada para utilizar software de reconocimiento facial que no sea en tiempo real con cámaras de vigilancia de circuito cerrado para seguir todos nuestros movimientos, sacando imágenes de las cuentas de redes sociales sin el consentimiento de la gente*. (…) *Siguen existiendo importantes lagunas en la prohibición del uso privado del reconocimiento facial y muchas otras formas de vigilancia biométrica remota*." Y, añade (aunque, al menos por el momento, no resultaría de aplicación en el caso español, en virtud de lo expuesto con anterioridad): "*La propuesta también sigue permitiendo el uso del reconocimiento facial en tiempo real de personas que sean sospechosas de entrar irregularmente en un Estado o vivir en él, algo que, sin ninguna duda, puede utilizarse como arma contra personas migrantes y refugiadas.*"[675]

[675] Amnistía Internacional, 2021.

Es interesante, en relación con ello, con carácter genérico, hacer alusión a la encomiable labor que realiza Interpol (que cuenta con el denominado "*Interpol Facial Recognition System*" -IFRC-, que contiene imágenes faciales remitidas por más de 160 países que crean una de las mayores bases de datos criminales del mundo) en aras de promover un uso eficaz pero responsable del uso de las tecnologías de reconocimiento biométrico, especialmente para fines de seguridad.

Y es que, por un lado, Interpol organiza de forma bianual el denominado Simposio Internacional sobre Dactiloscopia y Reconocimiento Facial, con la finalidad de que expertos de todo el mundo intercambien impresiones y compartan las últimas innovaciones a la vez que discuten sobre buenas prácticas; y, por otro lado, organiza dos veces al año reuniones del Grupo de Trabajo de Expertos en Reconocimiento Facial, un grupo consultivo sobre nuevas tecnologías que, entre otras, ha elaborado una guía de buenas prácticas relativas a la calidad, el formato y la transmisión de imágenes faciales para conseguir un reconocimiento preciso y efectivo. Además, con el mismo objetivo de mantenerse a la cabeza de un uso policial de las técnicas de reconocimiento biométrico innovador pero respetuoso con los derechos humanos, Interpol cuenta con su propio Reglamento sobre el Tratamiento de Datos[676], aplicable a toda operación de tratamiento de datos personales efectuada en el Sistema de Información de dicha organización.

Por su parte, de forma específica, en algunos países miembros ya se está poniendo de relieve la necesidad de regular de forma nacional algunas de las herramientas analizadas. Así, por ejemplo, en julio de 2019 Francia publicó un informe parlamentario en el que anunciaba la voluntad de regular e iniciar pruebas con sistemas de reconocimiento facial, estableciendo: "*Parece necesario desarrollar un marco legislativo de experimentación con el fin de probar estos sistemas en condiciones reales y asegurar nuestra soberanía, para que no seamos dependientes de las soluciones desarrolladas por los gigantes tecnológicos, y establecer un marco regulatorio que mejor se adapte a los usos*."[677] Y es que en tal documento se puso de manifiesto que, tras el experimento realizado durante la 135ª edición del Carnaval de la ciudad de Niza en febrero de ese año[678] bajo el marco legal del RGPD, se había advertido que la

[676] Véase Interpol, 2011.
[677] Baichère, 2019.
[678] Se realizaron pruebas diversas sobre individuos que habían ofrecido previamente su consentimiento búsqueda de niños perdidos, controles de flujos en los puntos de acceso, control de acceso restringido, etc..

regulación europea sobre técnicas de reconocimiento facial era insuficiente y que requería de una suplementación (especialmente respecto de su uso por parte de autoridades para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales o de ejecución de sanciones penales, incluidas la protección y prevención frente a las amenazas contra la seguridad pública).

Y es que, de acuerdo con lo expuesto, tales habilitaciones legales, si bien al parecer se llevarán a cabo de forma unitaria por la UE, tal y como ya se ha anunciado, pueden ser asimismo realizadas por los Estados Miembros, y ello podría llevarse a cabo de dos formas: o bien de modo flexible y abierto, o bien modo estricto y cerrado. Me explico.

Así, por un lado, respecto del tratamiento de datos biométricos destinados de forma unívoca a la identificación de una persona con fines de investigación criminal, el legislador podría optar por establecer, al igual que hizo respecto de la adopción de medidas tecnológicas de investigación (artículos 588bis LECrim y siguientes), unos principios rectores (a saber, especialidad, idoneidad, excepcionalidad, necesidad y proporcionalidad de la medida) y unos mínimos legales aplicables (por ejemplo, audiencia del Ministerio Fiscal, resolución judicial motivada, secreto de las actuaciones, etc), y a su vez prever una serie de medidas aplicables ("*numerus clausus*"), concediendo al juez cierta discrecionalidad para autorizar o no su aplicación, debiendo valorar de forma concreta las circunstacias de cada caso (lo cual aplica ya, por ejemplo, respecto de la obtención y análisis forzoso y tradicional -sin emplear IA-, con fines de investigación criminal, de las huellas dactilares y del ADN, datos biométricos que permiten la identificación del individuo, y que resultan legalmente previstas).

Respecto de las huellas dactilares, vemos a diario cómo los cuerpos policiales de toda España recogen los datos lofoscópicos de aquellas personas que resultan identificadas por la posible comisión de un delito, con base en lo dispuesto en los artículos 282 y siguientes de la LECrim. Y es que la averiguación o comprobación de la identidad del presunto delincuente entra dentro de las obligaciones y prerrogativas que la ley encomienda a la Policía Judicial, por lo que la captación de sus huellas dactilares, a efectos identificativos, resulta absolutamente imprescindible y no requiere de autorización judicial. Tales datos biométricos, además, quedan luego almacenados en una base de datos policial que será consultada en el futuro si resulta necesario, por ejemplo, a los efectos de comprobar la participación de la persona

detenida en el crimen investigado, mediante el cotejo de sus huellas dactilares con las halladas en el lugar de los hechos.

Respecto del ADN, sin embargo, el artículo 363 de la LECrim, tal y como ya se ha expuesto con anterioridad, establece la necesidad de autorizacion judicial para la extracción y análisis de las muestras biológicas de aquellos sospechosos de haber cometido un delito que no presten su consentimiento expreso para ello.

Así, a diferencia de lo que ocurre con las huellas dactilares, la obtención y el análisis forzoso del ADN de una persona para fines investigativos requiere de autorización judicial debidamente motivada, basada en los principios de proporcionalidad y razonabilidad. Y en aquellos casos en que se cuente con tal resolución judicial, su recogida y examen resultan absolutamente legales y constitucionales, puesto que se entiende que el juez se ha encargado de valorar los derechos fundamentales y los intereses en juego, y analizar las circunstancias del caso, y ha entendido que prevalece la necesidad de obtener y examinar los datos genéticos del individuo en cuestión, por lo que no se considera que se vulneren sus derechos a no declarar contra sí mismo, a no confesarse culpable y a la defensa, ya que la persona investigada puede igualmente negarse a prestar declaración judicial y emplear todos los medios a su alcance para rebatir los resultados de las pruebas de ADN, mediante la aportación de dictámenes periciales de parte, entre otros, para contradecir los informes del laboratorio oficial, por lo que tales derechos permanecen intactos.

Por otro lado, sin embargo, el legislador podría optar por fijar de forma rígida y precisa cuáles serían aquellos casos que permitirían el tratamiento de datos biométricos con fines de investigación criminal, a los efectos de no dejar margen de actuación alguna ni a las autoridades policiales ni judiciales, entendiendo bastante el análisis previo realizado por el poder legislativo respecto de la viabilidad jurídica de la medida.

Bajo mi punto de vista, no obstante, siempre resulta más útil y efectivo establecer una legislación más genérica y abierta y dejar cierto margen de flexibilidad para su adaptación al caso concreto, ya que luego la casuística es infinita y, especialmente en aquellos casos en que medie control judicial, no resulta necesario limitar tanto las opciones, puesto que ello puede causar, en ocasiones, el efecto contrario al que se pretende conseguir, por lo que yo sin duda

optaría por una regulación de "mínimos" similar a la actualmente establecida para las medidas de investigación tecnológicas en los mencionados artículos 588bis y siguientes de la LECrim.

A.4. *Riesgos jurídicos generales*

a) Posible falta de precisión y potencial discriminatorio

Una vez analizada la legislación actualmente aplicable a los sistemas de IA analizados, en primer lugar procede poner de manifiesto que, hoy en día, si bien existen *software* cada vez más potentes, sofisticados y avanzados, los sistemas de reconocimiento biométrico no cuentan con un nivel de infalibilidad y precisión tan elevado que permita garantizar la fiabilidad de sus resultados.

Así, tal y como dispone la Agencia Española de Protección de Datos: "*A diferencia de los procesos basados en contraseñas o certificados, que son 100% precisos (p. ej. una clave puede ser correcta o no serlo), la identificación o autenticación biométrica se basa en probabilidades (p. ej. una huella digitalizada proporcionará una correspondencia al 96% con un individuo). Existe una determinada tasa de falsos positivos (da por buena una suplantación) y falsos negativos (rechaza a un individuo autorizado)*."[679]

Los *software* de reconocimiento biométrico, por lo general, como ya se ha dicho, funcionan con técnicas de *Machine Learning* y *Deep Learning* que, por su naturaleza, requieren de ingentes cantidades de datos para poder entrenarse y conseguir así ir mejorando con el tiempo. La clave, no obstante, tal y como ya se ha advertido con anterioridad, está fundamentalmente en la calidad de los datos de que se nutren tales sistemas, no solo porque deben tener un origen lícito, sino también porque deben ser lo suficientemente respresentativos de la diversidad como para permitir al sistema detectar con igual (o, al menos similar) precisión los datos biométricos de todas las personas con independencia de su sexo, edad, raza o rasgos especiales, puesto que, en caso contrario, no solo podrían llevar a equívocos con nefastas consecuencias, sino también a tratos discriminatorios. Y es que los sistemas de IA solo aprenden de los datos que reciben, por lo que si un algoritmo solo se entrena con caras

[679] Y es que, por ejemplo, entre otros, ha quedado demostrado que el parecido biométrico entre familiares con rasgos físicos similares ha causado confusión a los sistemas de reconocimiento facial. Véase Ortiz, 2017.

blancas, únicamente sabrá cómo identificar con precisión los rostros de tal color[680], lo que da lugar al conocido fenómeno "*Garbage in, Garbage out*" (GIGO)[681].

No obstante, la clave del éxito no solo reside en la calidad de los datos, sino también en la calidad de los sistemas en sí mismos, siendo que estos deben tener una capacidad técnica suficiente para detectar los rasgos humanos incluso en condiciones que no sean las más deseables, a saber: en el caso del reconocimiento facial, escasa iluminación, presencia de elementos externos tales como gafas, gorra, mascarilla, etc; en el caso de las huellas dactilares, muestras incompletas, con polvo, etc; en el caso del ADN, muestras con poco contenido genético o con perfiles genéticos de más de una persona; en el caso del reconocimiento de voz, ruido de fondo, afonía, etc; y en el caso del reconocimiento de firma y escritura, imágenes borrosas, sucias, etc. Y ello solo se consigue, desde luego, mediante el entrenamiento del *software* con muestras diversas que contengan tales elementos pero, fundamentalmente, con la utilización de las tecnologías más punteras y sofisticadas, que si bien tienen un coste elevado, cada vez resultan más infalibles. En cualquier caso, la calidad de los sistemas tiene que ir revisándose, actualizándose y mejorándose de forma recurrente, tal y como se prevé en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA (Ley de IA) y se modifican determinados actos legislativos de la Unión, habida cuenta de los innumerables intentos de sabotaje que, con toda probabilidad, serán usados por los delincuentes, tales como, por ejemplo, los denominados "sistemas adversarios", diseñados de forma específica para tratar de burlar la precisión de los sistemas de reconocimiento facial.[682]

Respecto de ello, es interesante traer a colación los hallazgos efectuados por diversos estudios, especialmente en el ámbito del reconocimiento facial, que ponen en evidencia sus fallos de precisión y su potencial discriminatorio.

[680] Así, por ejemplo, los nativos americanos tienen la tasa de coincidencia falsa más alta de todas las etnias, ya que prácticamente no están representados en todos los conjuntos de datos de entrenamiento.
[681] En español, "basura adentro, basura afuera", lo que implica que un ordenador simplemente opera con los datos que se le proporcionan.
[681] Véase Pautov, Melnikov, Kaziakhmedov, Kireev & Petiushko, 2019.

Así, por un lado, Joy Buolamwini, investigadora del MIT Media Lab y autora de la tesis doctoral titulada "*Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers*" [683], al darse cuenta de que un sistema de reconocimiento facial detectaba las caras blancas de sus conocidos y no la suya, de raza negra (hasta que se puso una máscara blanca), decidió investigar si ello era un caso aislado o si respondía a un patrón de discriminación algorítmica que se repetía en los mencionados *software*.

Así, la mencionada investigadora decidió analizar los sistemas de reconocimiento facial de IBM, Microsoft y Face ++[684], que eran los que, aparentemente, contaban con los mayores niveles de precisión del mercado. No obstante, al experimentar con todos ellos, detectó que los tres funcionaban peor cuando se trataba de reconocer rostros de mujeres negras (y cuanto más negra era la piel, peores los resultados) y, asimismo, advirtió que, en concreto, IBM y Microsoft funcionaban mejor para reconocer las caras blancas masculinas, y, sin embargo, Face ++ tenía mejor precisión al detectar las caras negras masculinas. De los tres sistemas, además, descubrió que el de IBM era el que tenía un mayor margen de error de precisión, al marcar una diferencia del 34% entre la exactitud en la detección de los rostros de hombres blancos y de las mujeres negras, lo cual se reputa inaceptable (especialmente si pensamos en emplear el sistema para usos de investigación penal).

Por otro lado, también resulta significante el análisis que la organización American Civil Rights Union (ACLU) realizó en 2018 del sistema de reconocimiento facial de Amazon denominado "*Rekognition*". Y es que a través de las pruebas efectuadas se detectó que el mencionado *software* identificó, de forma incorrecta, a nada más y nada menos que veintiocho miembros del Congreso de EEUU como indivuduos fichados como delincuentes, lo cual resulta, cuanto menos, inquietante. Además, se advirtió que dichos errores se dieron de forma desproporcionada con personas de color, siendo que estas representan un 20% de los miembros del Congreso y, sin embargo, las falsas coincidencias entre ellas supusieron el 39% del total de los fallos del sistema (es decir, casi el doble).[685]

[683] Véase Buolamwini, s.f..
[684] De la compañía MEGVii, con acceso a una de las mayores bases de datos de imágenes de rostros chinos.
[685] Véase Snow, 2018.

En el año 2018, asimismo, la organización Big Brother Watch publicó un informe titulado "*Policing for the Future inquiry*"[686] que puso de manifiesto que la policía de Reino Unido estaba empleando el reconocimiento facial automatizado en espacios y eventos públicos sin ningún fundamento legal, lo cual resultaba altamente peligroso habida cuenta de que tal tecnología era extremadamente inexacta, habiendo llegado a identificar erróneamente a personas inocentes hasta en el 98% de los casos, existiendo un promedio del 95% de personas identificadas de forma equivocada.[687] Además, se puso de relieve que la tecnología utilizada por la policía no había sido probada para detectar posibles sesgos, a pesar de la existencia de múltiples estudios que advertían de la posibilidad de que los sistemas de reconocimiento facial identificaran erróneamente a mujeres y personas de color en niveles desproporcionados.

Por su parte, en el año 2019 la Policía Metropolitana de Londres hizo público el uso de cámaras de reconocimiento facial en tiempo real por las calles de la capital británica y anunció que estas iban a estar conectadas con bases de datos nutridas por listados de personas sospechosas de haber cometido delitos graves y violentos. En relación con ello, si bien un informe independiente realizado por el Human Rights Centre de la Universidad de Essex[688] denunció que la mayoría de las alertas generadas por las cámaras eran erróneas y que el sistema podía contener sesgos, desde la policía se afirmó que, tras someter a puebas al sistema durante meses, el 70% de los sospechosos buscados había sido identificado al pasar frente a las cámaras y, sin embargo, solo una de cada mil personas generó una falsa alarma.[689]

Asimismo, un informe sobre herramientas de reconocimiento facial publicado el 19 de diciembre de 2019 por el National Institute of Standards and Technology[690] de EEUU, bajo el título "*Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*"[691] puso de manifiesto que, si bien los algoritmos de mejor calidad son muy precisos y producen muchos menos errores que el resto, con carácter general podía concluirse que tales sistemas contienen fallos y que existe un diferencial mayor en los falsos positivos que en los falsos negativos (en

[686] Big Brother Watch, 2018.
[687] Asimismo, el estudio reveló que la policía almacenaba imágenes de personas inocentes identificadas erróneamente durante un mínimo de treinta días sin informarlas de tal circunstancia.
[688] Fussey & Murray, 2019.
[689] Véase BBC News, 2020.
[690] Dependiente del Departamento de Comercio del Gobierno de EEUU.
[691] Grother, Ngan & Hanaoka, 2019.

ocasiones, hasta cien veces más) respecto de los distintos grupos demográficos definidos por sexo, edad y raza o país de nacimiento.

Así, dicho estudio pone de relieve que, por ejemplo, las tasas de falsos positivos son hasta cien veces más altas en personas de África occidental y oriental y de Asia oriental, y más bajas en personas de Europa oriental.[692] Asimismo, concluyó que con las imágenes empleadas por las policías de EEUU, los falsos positivos más altos se daban respecto de los indios americanos y, finalmente, se advirtió de que los falsos positivos eran más altos en mujeres que en hombres, y en ancianos y en niños que en adultos de mediana edad.

Y la ya mencionada organización ACLU, además, por un lado, el 12 de marzo de 2020, junto con la organización New York Civil Liberties Union interpuso una demanda[693] frente al Departamento de Seguridad Nacional de EEUU, la Oficina de Aduanas y Protección Fronteriza de dicho país, el Servicio de Inmigración y Control de Aduanas de EEUU y la Administración de Seguridad del Transporte con el fin de que cesaran en su secretismo respecto del uso de sistemas de reconocimiento facial en los aeropuertos estadounidenses e hicieran públicos, entre otros, los contratos gubernamentales suscritos con aerolíneas, aeropuertos y otras entidades relacionadas con el uso de tal tecnología, las políticas y los procedimientos relacionados con la adquisición, el procesamiento y la retención de la información biométrica y el análisis de la efectividad de la tecnología de reconocimiento facial, por la posible existencia de errores y sesgos. [694]

Y, por otro lado, el 24 de junio de 2020 tal organización hizo pública[695] la historia de Robert Julian-Borchak Williams, protagonista del primer caso (al menos, conocido) de detención policial errónea en EEUU como consecuencia del uso de un sistema de reconocimiento facial. Y es que el mencionado ciudadano americano, de raza negra y residente en Detroit (Michigan, EEUU), fue arrestado en el jardín de su casa ante la aterrorizada mirada de su esposa y sus dos hijas menores porque, al parecer, el sistema de reconocimiento facial de la Policía de Michigan (que introdujo en su base de datos de licencias de conducir las imágenes captadas

[692] Aunque, con una serie de algoritmos desarrollados en China, tal efecto se invierte, con tasas bajas de falsos positivos en las caras de Asia oriental.
[693] Véase American Civil Liberties Union, 2020.
[694] American Civil Liberties Union, 2020.
[695] American Civil Liberties Union, 2020.

por la cámara de seguridad de una tienda de Detroit que mostraba a una persona sustrayendo relojes), le identificó como el autor del robo. Tras ello, el mencionado ciudadano pasó una noche en el calabozo, y se recogió por la policía muestra de sus huellas digitales y ADN, si bien, al día siguiente, un agente de la Policía de Detroit reconoció que el sistema de IA de la Policía de Michigan, adquirido a la compañía Data Works Plus, se había equivocado, por lo que Robert Williams fue liberado, con todas las nocivas consecuencias que el terrible error conllevó para él y su familia.

En relación con tales controvertidos sistemas, justamente, en el año 2020 la propia compañía Amazon prohibió a la policía el uso de la herramienta de reconocimiento facial comercializado por ella ("*Rekognition*") durante un año (aunque actualmente la moratoria tiene carácter indefinido), a raíz del incidente ocurrido con el ciudadano George Floyd, que dio lugar a protestas masivas en EEUU bajo el denominado movimiento "*Black Lives Matter*". Y es que tal gigante tecnológico, tras recibir numerosas críticas por parte de múltiples defensores de derechos humanos por el posible sesgo racial de su sistema, decidió hacer un parón en el uso policial del mismo "*hasta que los gobiernos establezcan regulaciones más estrictas para regular el uso ético de la tecnología de reconocimiento facial*", tal y como manifestó en el comunicado publicado el 10 de junio de 2020.[696]

En el mismo sentido y con igual fundamento, el 8 de junio de 2020 la compañía IBM también anunció, mediante una carta remitida por su Consejero Delegado, Arvind Krishna, al Congreso de EEUU, que dejaba de ofrecer su *software* de reconocimiento facial para "*la vigilancia masiva, la discriminación por perfil racial, las violaciones de los derechos humanos y las libertades básicas*"[697] y le urgió a regular el uso de tal tecnología, manifestando que era "*el momento de iniciar un diálogo nacional sobre si los cuerpos policiales deben emplear la tecnología de reconocimiento facial y de qué manera*".[698]

Y, asimismo, Microsoft, el 11 de junio de 2020, en una conversación mantenida por su Presidente, Brad Smith, con el periódico The Washington Post, hizo pública su decisión de no vender sus sistemas de reconocimiento facial a los departamentos de policía del país hasta

[696] About Amazon, s.f..
[697] IBM, s.f..
[698] *Idem.*

que su uso, con garantías de respeto a los derechos humanos, estuviera regulado por una ley federal.[699]

En relación con ello, el 25 de junio de 2020 los Senadores de EEUU Ed Markey y Jeff Merkley presentaron un proyecto de ley denominado "*Facial Recognition and Biometric Technology Moratorium Act*"[700] [701] con la finalidad de prohibir la vigilancia biométrica (inlcuyendo reconocimiento facial, reconocimiento de voz y de otras características físicas inmutables) por parte del Gobierno Federal estadounidense -y sus agencias de policía- sin autorización legal explícita, y limitar la concesión de ciertos subsidios federales de seguridad pública únicamente a aquellos gobiernos estatales y locales que desarrollaran sus propias moratorias. En relación con ello, el Senador Markey manifestó "*La tecnología de reconocimiento facial no solo representa una grave amenaza para nuestra privacidad, sino que pone en peligro físicamente a los afroamericanos y otras poblaciones minoritarias en nuestro país* (…) *En un momento en que los estadounidenses exigen que abordemos el racismo sistémico en la aplicación de la ley, el uso de la tecnología de reconocimiento facial es un paso en la dirección equivocada. Los estudios demuestran que esta tecnología trae discriminación racial y prejuicios*".[702]

Y es que si bien en febrero de ese mismo año el propio Senador estadounidense Ed Markey, junto con el Senador Cory Booker, ya presentó un proyecto de ley bajo el título "*The Etichal Use of Facial Recognition Act*"[703], que también tenía como finalidad establecer una moratoria sobre el uso de los sistemas de reconocimiento facial por parte del Gobierno federal pero contemplaba diversas excepciones, tales como su utilización por parte de las fuerzas policiales con autorización judicial[704], en este último proyecto de ley, especialmente como consecuencia de los lamentables hechos ocurridos con el mencionado ciudadano George Floyd, el contenido legislativo fue más limitante, tajante y firme que en el anterior.

[699] Véase Greene, 2020.
[700] Véase Markey & Merkley, 2020.
[701] Tal proyecto de ley está respaldado por ACLU, Fight for the Future, Color of Change, MediaJustice, MPower Change, Electronic Frontier Foundation, Electronic Privacy Information Center (EPIC), Athena Coalition, Jewish Voice for Peace, Free Press, Project on Government Oversight, "*Center on Privacy and Technology*" de la Universidad de Georgetown y el Open Technology Institute-New America.
[702] Ed Markey, 2020.
[703] Véase Jeff Merkley, 2020.
[704] Véase Solon, 2020.

No obstante, las voces partidarias del uso de tales sistemas de IA de reconocimiento facial no se han hecho esperar, siendo que entienden que su prohibición total no es la solución, ya que lo realmente necesario es que los gobiernos desarrollen un marco legal basado en la evidencia y regulen de forma estricta la calidad y las aplicaciones específicas de tales tecnologías, habida cuenta de que los algoritmos, empleados de forma responsable, pueden servir justamente para corregir el actual sistema sesgado y defectuoso.[705]

Así, en relación con todo lo anteriormente expuesto, procede poner de manifiesto que, con el uso de los sistemas de reconocimiento de datos biométricos existentes en la actualidad (especialmente, con los de reconocimiento facial), existe un peligro real de vulneración de derechos fundamentales de los ciudadanos.

Y es que, por un lado, habida cuenta de las expuestas elevadas tasas de error de los mencionados sistemas (que varían, no obstante, tal y como se ha visto, en función de la calidad de los mismos y de los datos empleados por estos), entiendo que su uso en el ámbito de la investigación criminal podría llegar a suponer una grave infracción del derecho a libertad (artículo 6 de la Carta de Derechos Fundamentales de la Unión Europea y artículo 17 de la Constitución Española) y a la presunción de inocencia (artículo 48 de la Carta de Derechos Fundamentales de la Unión Europea artículo 24 de la Constitución Española) de los ciudadanos, tal y como se puso de manifiesto, por ejemplo, en el mencionado caso de Robert Julian-Borchak Williams, que fue detenido y pasó una noche en el calabozo a pesar de ser inocente, únicamente con base en los resultados de un *software* de reconocimiento facial empleado por la policía que resultó estar equivocado.

Además, en caso de que un individuo fuera erróneamente arrestado, durante la celebración de un evento de culto religioso (por ejemplo, durante la celebración de una misa) o de legítima protesta pública (por ejemplo, durante la marcha de una manifestación pacífica), como consecuencia de la equivocación de un sistema de reconocimiento facial, podrían asimismo quedar vulnerados tanto su derecho a la libertad de culto (artículo 10 de la Carta de Derechos Fundamentales de la Unión Europea y artículo 16 de la Constitución Española), su derecho de reunión (artículo 12 de la Carta de Derechos Fundamentales de la Unión Europea y artículo

[705] Véase Rao, 2020.

21 de la Constitución Española) y su derecho a la libertad de expresión (artículo 11 de la Carta de Derechos Fundamentales de la Unión Europea y artículo 20 de la Constitución Española), respectivamente.

Por otro lado, debido a las grandes diferencias que, en muchas ocasiones, se dan entre los niveles de precisión de los sistemas de reconocimiento facial a la hora de detectar de forma correcta los datos biométricos de unos grupos demográficos y de otros (por lo general con tendencia a los sesgos en contra de las minorías, como se ha visto), considero que su uso en el ámbito de la investigación criminal podría llegar a suponer un serio quebranto del derecho a la igualdad y a la no discriminación (artículos 20, 21 y 23 de la Carta de Derechos Fundamentales de la Unión Europea y artículo 14 de la Constitución Española) de los ciudadanos, tal y como, por ejemplo, también se puso de manifiesto en el antedicho caso de Robert Julian-Borchak Williams, ciudadano estadounidense erróneamente arrestado, justamente de raza negra.

Y, además, finalmente, en aquellos casos en que un individuo fuera arrestado en público como consecuencia del erróneo resultado otorgado por un sistema de reconocimiento facial, sin duda podría verse vulnerado su derecho al honor (artículo 18 de la Constitución Española) e incluso su derecho a la dignidad (artículo 1 de la Carta de Derechos Fundamentales de la Unión Europea y artículo 10 de la Constitución Española), así como, en los casos más graves, su derecho a la integridad física y moral y la de sus familiares (artículo 3 de la Carta de Derechos Fundamentales de la Unión Europea y artículo 15 de la Constitución Española), lo cual también se puso de manifiesto en el expuesto caso de Robert Julian-Borchak Williams, habida cuenta de la repercusión que la noticia causó en su vecindario y, sobre todo, del impacto emocional que ocasionó en él y en su familia, que tuvo que presenciar cómo era detenido por la policía.

En cuanto a los sistemas de reconocimiento de huellas dactilares y de ADN, si bien son conocidos por ser los métodos más infalibles que hoy en día existen, procede poner de manifiesto que su éxito, en caso del empleo de IA, depende, como ocurre con todos los sistemas de tal tecnología que emplean datos biométricos, no solo de la calidad de los algoritmos, sino también de la calidad de los datos en ellos contenidos.

Así, en relación con la precisión de tales herramientas, hay que decir que esta depende, fundamentalmente (al igual que sucede en con los sistemas de reconocimiento de voz y de firma y escritura), de la calidad de las muestras, que no siempre resulta ser óptima y adecuada.

Y es que, por un lado, especialmente cuando se trata de muestras dubitadas halladas en el lugar de los hechos o de muestras captadas por ciudadanos legos en las respectivas técnicas de captación, estas pueden contener elementos perturbadores tales como, entre otros, suciedad, mezcla con otras muestras, trozos incompletos, ruidos de fondo, etc, que pueden conllevar dificultades añadidas y mermar el éxito de los mencionados sistemas, con las consecuencias que ello puede conllevar, tal y como ya se ha avanzado con anterioridad.

Y, por otro lado, tales muestras pueden resultar muy limitadas o reducidas, de modo que no reflejen una imagen fiel de la realidad, lo que podría llevar a perpetuar los patrones discriminatorios que desafortunadamente se han venido llevando a cabo de forma histórica. Así, por ejemplo, si las bases de datos de los sistemas de IA de reconocimiento de huellas dactilares o ADN están nutridas únicamente con las muestras dactilares o genéticas de aquellas personas que han resultado detenidas y condenadas a lo largo de los tiempos, y estas han sido, en su gran mayoría, de raza negra (como consecuencia de prácticas policiales discriminatorias), los resultados siempre van a favorecer a los presuntos delincuentes de raza blanca, que gozarán de mayor impunidad que los de raza oscura, lo cual es absolutamente inadmisible. Es por ello que, en mi opinión, una buena solución sería la de dar a tales sistemas acceso, de forma limitada y con control judicial, a la información de bases de datos más generalizadas, tales como las del DNI o pasaporte, con el fin de asegurar que las muestras son representativas del conjunto de la sociedad y democratizar y objetivizar así los resultados de los mismos.

En el ámbito del reconocimiento de voz, especialmente relevante es centrar la atención en un potencial elemento discriminatorio que debe tenerse en cuenta para tender a su eliminación: el del mayor éxito en relación con las lenguas mayoritarias en detrimento de las minoritarias. Y es que, principalmente por cuestiones económicas, los algoritmos de los sistemas de reconocimiento de voz resultan entrenados para funcionar con idiomas de habla mayoritaria, de modo que funcionan con mayor efectividad en tales casos, lo que supone una clara discriminación frente a aquellas personas con lenguas de habla minoritaria.

Así, una víctima de violencia de género de origen pakistaní que hable un dialecto del urdu y necesite hacer uso de un *chatbot* para poder solicitar ayuda policial en caso de que su ex pareja incumpla las medidas cautelares de alejamiento impuestas, por ejemplo, seguramente no tendrá el mismo acceso que el que tendría una víctima de origen francés, lo cual resulta inadmisible y debe tender a solucionarse de la forma más eficaz y rápida posible, habida cuenta de los derechos fundamentales que están en juego, aunque ello pueda considerase antieconómico.

Y no solo eso, ya que la mencionada discriminación frente a las minorías se da incluso entre los distintos acentos existentes en las lenguas mayoritarias. Así, un estudio publicado en la revista estadounidense Proceedings of the National Academy of Sciences USA[706] demostró que los sistemas de IA de reconocimiento de voz están sesgados contra los hablantes de raza negra, habiendo llegado a la conclusión de que los programas de las principales empresas tecnológicas, incluidas Apple y Microsoft, tenían aproximadamente el doble de probabilidades de transcribir incorrectamente los audios emitidos por personas de raza negra, con su singular acento en inglés, que los emitidos por personas de raza blanca.

En relación con ello, y con el fin de evitar las negativas consecuencias que la mala calidad de los datos puede ocasionar, tal y como ya se ha expuesto con anterioridad, la Propuesta de Reglamento de IA de la Comisión Europea establece una serie de obligaciones y requisitos para los sistemas de IA calificados de alto riesgo, con el fin de reforzar las garantías en tal sentido.[707]

Así, en el artículo 10 se establecen unos estándares de entrenamiento, validación y pruebas de datos, disponiendo el apartado 2 que tales sistemas estarán sujetos a prácticas adecuadas de gestión y gobernanza, que afectarán, entre otras, al examen de los posibles sesgos y a la identificación de posibles lagunas o deficiencias de datos, así como a la forma de abordarlas. Además, se dispone en el apartado 3 que "*Los conjuntos de datos de entrenamiento, validación y prueba serán pertinentes y representativos, carecerán de errores y estarán completos. Asimismo, tendrán las propiedades estadísticas adecuadas, también en lo que respecta a las personas o los grupos de personas en relación con los que se pretenda utilizar*

[706] Koenecke & otros, 2020.
[707] Véanse págs. 192-197.

el sistema de IA de alto riesgo, cuando proceda. Los conjuntos de datos podrán reunir estas características individualmente para cada dato o para una combinación de estos."

Ello, sin duda es una buena noticia, aunque habrá que ver cómo queda finalmente redactado el texto definitivo para valorar la suficiencia y la utilidad de las medidas establecidas.

En el ámbito del reconocimiento de emociones, existe una buena parte de la doctrina que se muestra escéptica sobre la precisión de sus resultados y, por ende, reacia a su uso. En concreto, merece especial mención en tal sentido el artículo "*Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements*"[708], publicado el 17 de julio de 2019 en el portal de Sage Journals y que cuestiona si realmente puede inferirse de modo razonable el estado emocional de una persona a partir de sus movimientos faciales.

En relación con todo ello, en ocasiones me surgen dudas y, entre otras cuestiones, con el fin de mejorar y objetivar la calidad de los datos empleados en el seno de las investigaciones penales, en ocasiones me planteo: ¿por qué no se recaban los datos biométricos de los ciudadanos en el momento de expedir o renovar su carné de identidad o pasaporte? ¿acaso ello no aumentaría exponencialmente el tamaño de las bases de datos y facilitaría enormemente las tareas de identificación en las investigaciones criminales, por ejemplo?. La respuesta, no obstante, una vez expuesto lo anterior, no me resulta tan obvia.

Y es que, como en todas las actuaciones que pueden implicar alguna limitación de derechos fundamentales, antes de realizar cualquier propuesta procede analizar si la medida resultaría útil, idónea, necesaria y proporcionada.

Por un lado, sin duda, el hecho de recopilar de forma masiva los datos biométricos de todos los ciudadanos en el momento de la expedición y/o renovación de su DNI o pasaporte permitiría aumentar de forma exponencial, para fines de investigación criminal, el contenido de las bases de datos nacionales e internacionales, que actualmente básicamente están nutridas por datos biométricos de las personas investigadas o condenadas por la comisión de delitos. Y es que, el hecho de poder "lanzar contra la base de datos" cualquier información biométrica

[708] Feldman, Adolphs, Marsella, Martinez & Pollack, 2019, págs. 1-68.

de posible interés criminal y saber que, prácticamente con seguridad, ello implicará hallar una coincidencia de forma automática y lograr la identificación buscada, es un auténtico lujo. Lamentablemente, sin embargo, no deja de ser una utopía.

Hoy en día, tal y como es sabido, la cesión de las huellas dactilares de los ciudadanos en el momento de expedición y/o renovación del DNI o pasaporte, con fines de identificación, está absolutamente normalizada (a pesar de que los sectores más radicales en materia de protección de datos personales se han mostrado contrarios a ello), ya que se entiende una medida idónea, necesaria y proporcionada, habida cuenta de que los gobiernos deben contar con infrmación precisa[709] y actualizada sobre sus nacionales, a efectos simplemente de identificar y tener un cierto orden y control de aquellos que están bajo su jurisdicción (necesarios para el buen funcionamiento del Estado de Derecho). No obstante, la cesión y biometrización masiva y obligatoria de datos tan sensibles como pueden ser las propias huellas dactilares, los rasgos faciales, el ADN o la voz de las personas, puede resultar excesiva y contraria a los derechos humanos.

En mi opinión, el problema no yace en el mero hecho de recopilar y tratar la antedicha información para fines que podrían resultar legítimos (a saber, la prevención o investigación criminal, por ejemplo), sino que este surge por la pérdida de control de tales datos personales que ello implicaría para los ciudadanos, con los potenciales perjuicios irreparables que su mal uso podría entrañar.

Y es que, por lejano que pueda parecer, las brechas de seguridad, se dan; las quiebras del Estado de Derecho, también; y los abusos de poder, ni hablemos, por lo que no resulta tan remota la posibilidad de que los datos biométricos cedidos cayeran en manos de personas o compañías con fines ilegítimos.

Ante tal planteamiento, no obstante, en ocasiones me pregunto si ser tan garantista con este tipo de actuaciones no hace más que anclarnos en el pasado e impedirnos avanzar, siendo que además vivimos en un mundo cada vez más digital en que, nos guste o no, en unos años -no demasiados-, la gran mayoría de las relaciones humanas dejarán huella tecnológica. Así, me cuestiono si es muy *naïf* ser tan puristas, desde el ámbito público, con las captaciones masivas

[709] Siendo las huellas dactilares uno de los elementos identificativos más exactos que existen.

de datos personales (en este caso, biométricos), puesto que sin duda, con un uso regulado -por supuesto-, podrían elevar la eficiencia de la prevención y la investigación criminal y, por ende, la seguridad pública, a niveles inimaginables. Y es que ¿qué sentido tendría que las compañías privadas tuvieran acceso masivo a los datos biométricos de los ciudadanos -los rasgos faciales y de voz cedidos al *smartphone*, la firma plasmada en un soporte capaz de biometrizar, la información genética cedida para realizar estudios, etc- y en cambio el Estado se mantuviera al margen, a pesar de los enormes beneficios que, para fines legítimos, le podrían reportar? La respuesta es simple: la clave está en el carácter obligatorio o voluntario de la cesión de los datos.

Así, no es lo mismo que una persona consienta que su *smartphone* emplee sus rasgos faciales para desbloquear el teléfono que que esta resulte obligada, por ley, a cederlos al Estado en el momento de su expedición y/o renovación del DNI o pasaporte. Ello, no obstante, me lleva a concluir que una solución intermedia y jurídicamente viable sería la de la cesión voluntaria de tales datos, de modo que cada ciudadano que acudiera a una comisaría a expedir o renovar su DNI o pasaporte pudiera, previa firma de un consentimiento informado (con explicación clara de lo que implicaría realmente la cesión de sus datos biométricos y sus potenciales riesgos), autorizar la captación y el tratamiento de sus datos biométricos para los fines legalmente establecidos (entre otros, la investigación criminal), con posibilidad de revocar tal consentimiento en cualquier momento (lo cual, no obstante, no es perfecto, ya que una vez se han cedido los datos, la realidad es que se ha perdido el control sobre ellos para siempre). No obstante, en tal caso, lo que resultaría fundamental es que hubiera una regulación legal clara de en qué supuestos, bajo qué circunstancias y con qué requisitos tales datos cedidos podrían ser tratados, en aras de que los ciudadanos pudieran conocer, de forma simple y transparente, cuáles serían las consecuencias de su decisión (previa autorización judicial, por ejemplo, para identificar cadáveres, investigar delitos graves, etc siempre con ponderación de los principios de especialidad, idoneidad, excepcionalidad, necesidad y proporcionalidad, del mismo modo que se exige *ex* artículo 588bis LECrim para la investigación con medios tecnológicos, etc).

Ello, sin duda, no tendría el mismo impacto en las bases de datos que el que resultaría de establecer una cesión forzosa de los mencionados datos biométricos en el momento de expedición y/o renovación del pasaporte o del DNI, pero sin duda ayudaría a mejorar la calidad de estas y de los sistemas de IA que emplean tal clase de datos. Y es que, sin duda,

cuantos más datos biométricos (y más variados) alimenten las bases de datos policiales y judiciales, mayor será su nivel de precisión (puesto que los algoritmos podrán ser entrenados con datos diversos) y menor será su nivel de discriminación (puesto que si únicamente se nutren de los datos biométricos de las personas investigadas o condenadas por la comisión de delitos, ello va a perpetuar las prácticas pasadas, en muchos casos cuestionables).

Es necesario poner de manifiesto, admás, en relación con ello, que hoy en día, y justamente ante el problema de la escasez de los datos que nutren las bases de datos policiales, ciertos cuerpos de policía están buscando soluciones de forma unilateral, lo cual resulta altamente peligroso y no hace más que dejar patentes los grandes beneficios que la propuesta que he expuesto podría reportar. Así, por ejemplo, la ONG ProPublica publicó en 2016 un informe que advertía de que las fuerzas policiales de Florida y otros estados de EEUU estaban creando bases de datos de ADN privadas, en parte mediante la recopilación de muestras ADN de personas no sospechosas de haber cometido ningún delito, cedidas de forma "voluntaria" (incluso de menores, sin consentimiento parental ni, en su defecto, autorización judicial).[710] Y en relación con ello, la organización ACLU San Diego en 2017 presentó una demanda frente a la ciudad de San Diego con el fin de poner fin a las prácticas del Departamento de Policía de tal urbe, que recolectaba muestras de ADN de menores en los términos descritos.[711]

Por su parte, según denunció en 2019 la organización ACLU Massachussets, el FBI recopila datos sobre los rostros, iris, formas de caminar y voces de los ciudadanos, lo que permite al gobierno estadounidense identificarlos, rastrearlos y monitorizarlos de manera generalizada (de hecho, según la referida ONG, el FBI puede comparar las caras de los ciudadanos estadounidenses con al menos seiscientos cuarenta millones de imágenes de adultos que residen en tal país).[712]

No obstante, la solución, en la mayoría de ocasiones, no se halla en la captación masiva de datos, sino la inversión en medios materiales y personales para resolver los problemas estructurales o de fondo que acechan, en este caso, a la prevención y a la investigación de delitos. Y es que, sin duda, los gobiernos deberían plantearse tal cuestión como prioritaria y destinar los recursos necesarios para establecer mejoras en tal ámbito, sin recurrir a "lo fácil"

[710] Feldman, Adolphs, Marsella, Martinez & Pollack, 2019, págs. 1-68.
[711] Véase American Civil Liberties Union, 2017.
[712] Véase Crockford, 2019.

que es recopilar datos de forma indiscriminada y masiva y ejercer así un control total de la población a costa, no obstante, de vulnerar de forma frontal sus derechos y libertades. Un equilibrio en tal sentido, de captación de datos de forma prudente, transparente, proporcionada y, sobre todo, regulada, y de inversión de recursos, desde luego sería la opción más óptima y deseable, si bien entiendo que no resulta fácil de conseguir, pero no por ello debe descartarse.

b) Posible vulneración del derecho a la privacidad y a la protección de datos personales

En segundo lugar, es fundamental hacer referencia a los posibles riesgos jurídicos que giran en torno a la calidad y la legalidad de los datos empleados en lo relativo a su tratamiento, que como ya se ha dicho en páginas anteriores, es definido por el artículo 5.b) de la LO 7/21, de 26 de mayo, como "*cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción*."

Y es que la viabilidad y el éxito de los sistemas de reconocimiento de datos biométricos no solo depende de su precisión, sino que va mucho más allá, habida cuenta de que estos deben cumplir con la legislación vigente en materia de protección de datos para poder operar con todas las garantías.

Así, a modo de ejemplo, en el ámbito civil (no de investigación criminal), en 2015 se interpuso en el estado de Ilinois (EEUU) una demanda colectiva contra Facebook por la captación de los datos biométricos (y su posterior uso) de las fotografías de miles de usuarios sin su consentimiento. Y es que la Ley de protección de datos biométricos de dicho estado ("*Biometric Information Privacy Act 2008*"[713]), que es una de las más estrictas de EEUU, requiere que para que el consentimiento de los consumidores para la recolección y el uso de sus datos biométricos sea válido, la empresa detalle previamente qué información captará, cómo la Véase (American Civil Liberties Union, 2017) utilizará y durante cuánto tiempo la

[713] Véase Illinois General Assembly, s.f..

conservará. El caso, uno de los más sonados en este ámbito, especialmente por la elevada cifra en juego (seiscientos cincuenta millones de dólares), no obstante, fue finalmente resuelto mediante la aceptación por parte de Facebook de pagar tal cantidad a modo de indemnización a los múltiples usuarios afectados[714] [715], por lo que los tribunales no tuvieron que decidir.

En relación con ello, resulta interesante traer a colación una de las primeras sentencias que se han dictado en el ámbito europeo en relación a la legalidad de los sistemas de reconocimiento facial automatizado empleados por la policía con fines de investigación criminal, que justamente ha dotado de igual peso a la necesidad de que estos resulten precisos y no contengan sesgos, y a la necesidad de que estos respeten el derecho a la privacidad y la protección de los datos personales de los ciudadanos.

Así, el 11 de agosto del 2020 la Corte de Apelaciones de Inglaterra y Gales dictó una sentencia[716] que determinó que el uso por parte de la Policía del Sur de Gales de sistemas de reconocimiento facial automatizado en lugares y eventos públicos resultaba ilegítimo y vulneraba los derechos humanos.

En concreto, la herramienta de IA analizada (denominada "*AFR Locate*") procedía a comparar los datos biométricos de la imágenes captadas por las cámaras de seguridad de espacios públicos con los de las imágenes contenidas en las bases de datos policiales[717], con el fin de observar si existían o no coincidencias y localizar así a personas sospechosas.

Al examinar el mencionado sistema, el Tribunal británico determinó que su uso resultaba proporcionado, si bien entendió que no podía entenderse legal, habida cuenta de que, por un lado, otorgaba demasiada discreción a los agentes de policía para decidir dónde se podía emplear[718] y qué individuos debían ser incluidos en las bases de datos; que, por otro lado, no se había realizado una adecuada evaluación de los riesgos que su uso podía acarrear para los

[714] Véase Marotti, 2020.
[715] Véase Marotti, 2020.
[716] Véase más en Courts and Tribunals Judiciary, 2020.
[717] A pesar de que, en caso de no haber coincidencia alguna, las imágenes captadas se eliminaban automáticamente, de forma inmediata.
[718] Se puso de manifiesto, por ejemplo, que el uso de tales sistemas no quedó limitado solamente a aquellas áreas en las que razonablemente podría creerse que estaban presentes las personas incluidas en las bases de datos de sospechosos.

derechos y libertades de los ciudadanos, en concreto para el derecho a la privacidad reconocido en el artículo 8 de la Convención Europea de Derechos Humanos; y, finalmente, que la Policía del Sur de Gales no había llevado a cabo, con carácter previo a su utilización, pruebas y análisis suficientes para determinar si el *software* de reconocimiento facial empleado contenía sesgos, principalmente raciales o de género.[719]

En relación con ello, ya en 2018 la ONG Big Brother Watch publicó un estudio que advertía de que había alrededor de diecinueve millones de imágenes de detenidos en la Base de Datos Nacional de la Policía, y de estas, casi trece millones estaban biometrizadas y, por ende, eran susceptibles de ser identificadas por sistemas de IA de reconocimiento facial y, sin embargo, el Comisionado de Biometría había estimado que cientos de miles de ellas correspondían a personas inocentes.[720]

Así, en el ámbito de la UE, los sistemas de reconocimiento de datos biométricos no solo deben cumplir con estándares legales de precisión e igualdad y no discriminación sino también de privacidad y protección de datos. Y en este último sentido, en el ámbito de la investigación penal deben quedar sujetos, en caso de ser empleados por las autoridades, a lo dispuesto en la LO 7/21, de 26 de mayo, tal y como anteriormente ya se ha expuesto, siendo que en caso contrario podrían llevarse a cabo graves vulneraciones de derechos de los ciudadanos, que perderían el control de sus propios datos biométricos y, por ende, de su identidad.

Y es que el espíritu de la LO 7/21, de 26 de mayo (al igual que el del RGPD), tal y como se desprende de lo dispuesto en su Capítulo IV, es el de fomentar la responsabilidad proactiva de los responsables del tratamiento de datos personales, con orientación a la gestión del riesgo. Así, en concreto, el artículo 27.1 dispone con carácter general que:

"*El responsable del tratamiento, tomando en consideración la naturaleza, el ámbito, el contexto y los fines del tratamiento, así como los niveles de riesgo para los derechos y libertades de las personas físicas, aplicará las medidas técnicas y organizativas apropiadas para garantizar que el tratamiento se lleve a cabo de acuerdo con esta Ley Orgánica y con*

[719] Véase Courts and Tribunal Judiciary, 2020.
[720] Ferries, 2018.

lo previsto en la legislación sectorial y en sus normas de desarrollo. Tales medidas se revisarán y actualizarán cuando resulte necesario."

Y, por su parte, el artículo 35 establece la obligación específica de realizar una evaluación de impacto previa al mencionado tratamiento en los términos siguientes:

"*1. Cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, suponga por su naturaleza, alcance, contexto o fines, un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, con carácter previo, una evaluación del impacto de las operaciones de tratamiento previstas en la protección de datos personales.*

2. La evaluación incluirá, como mínimo, una descripción general de las operaciones de tratamiento previstas, una evaluación de riesgos para los derechos y libertades de los interesados, las medidas contempladas para hacer frente a estos peligros, así como las medidas de seguridad y mecanismos destinados a garantizar la protección de los datos personales y a demostrar su conformidad con esta Ley Orgánica. Esta evaluación tendrá en cuenta los derechos e intereses legítimos de los interesados y de las demás personas afectadas.

3. Las autoridades de protección de datos podrán establecer una lista de tratamientos que estén sujetos a la realización de una evaluación de impacto con arreglo a lo dispuesto en el apartado anterior y, del mismo modo, podrán establecer una lista de tratamientos que no estén sujetos a esta obligación. Ambas listas tendrán un carácter meramente orientativo."

Y, en concreto, en los supuestos de tratamiento de datos biométricos, entiendo que resultaría de aplicación lo dispuesto en el artículo 36 del antedicho cuerpo legal, que establece en su apartado 1 (con subrayado propio):

"*El responsable o el encargado del tratamiento consultará a la autoridad de protección de datos, antes de proceder al tratamiento de datos personales que vayan a formar parte de un nuevo fichero, en cualquiera de las siguientes circunstancias:*

a) Cuando la evaluación del impacto en la protección de los datos indique que el tratamiento entrañaría un alto nivel de riesgo, a falta de medidas adoptadas por el responsable para mitigar el riesgo o los posibles daños.

b) Cuando el tipo de tratamiento pueda generar un alto nivel de riesgo para los derechos y libertades de los interesados, en particular, cuando se usen tecnologías, mecanismos o procedimientos nuevos."

Así, el responsable del tratamiento de los datos biométricos en cada caso deberá tener muy claro, con carácter previo, cuál es su función, identificar qué tipo de procesos y actividades lleva a cabo y detectar los posibles riesgos que estos pueden implicar de forma integrada. Y es que la gestión de los riesgos en materia de protección de datos que debe llevar a cabo cada responsable a los efectos de legitimar el tratamiento de datos personales que pretende realizar, podría compararse con la gestión de riesgos en materia penal que el artículo 31bis del Código Penal exige a las personas jurídicas para eximirlas de responsabilidad criminal en el caso de comisión de delitos por parte de sus representantes legales o de sus empleados en los términos previstos en dicho precepto, ya que lo que se solicita legalmente en ambos casos es que los responsables detecten e identifiquen los riesgos (en materia de protección de datos y en materia penal, respectivamente), y los gestionen e intenten prevenirlos de la forma más adecuada y correcta posible.

En concreto, a la hora de llevar a cabo una adecuada y completa evaluación de riesgos, siempre tiene que contemplarse y preverse por el responsable del tratamiento el peor de los escenarios posibles: el de la variación sustancial de las circunstancias, que puede ocurrir de forma legítima (por cambio de dirección del gobierno, por ejemplo) o incluso ilegítima (por un ciberataque o un golpe de Estado, entre otros). Y es que ello puede tener un impacto devastador en el ámbito de la protección de datos biométricos, por lo que debe preverse de antemano.

Y es que, en relación con ello, la Administración Pública tiene la obligación de velar, especialmente, por los intereses de toda la sociedad. Así, a la hora de llevar a cabo una evaluación de riesgos en materia de tratamiento de datos biométricos, en concreto, en el ámbito de la justicia, por ejemplo, sería muy importante contemplar no solo la situación

política actual, sino también una eventual quiebra del sistema o del Estado de Derecho y preguntarse: ¿qué sucedería con esos datos personales que se tratan y que, por ende, están a disposición de la Administración, si hubiera un golpe de Estado? o, ¿qué podría llegar a ocurrir si se decreta un estado de excepción o hay una guerra y se decreta un estado de sitio? ¿y si se sufre un ciberataque por parte de un estado totalitario que no respeta los derechos humanos? ¿qué consecuencias tendría todo ello para los ciudadanos?, etc.

Y, asimismo, una buena posterior gestión de riesgos implicaría la realización de un análisis crítico por parte de los responsables del tratamiento de los datos y un estudio de las alternativas existentes a dicha operación. Así, por ejemplo, si la instalación de un sistema de IA de tratamiento automatizado de datos biométricos puede evitarse mediante la contratación de más personal y ello supone la misma inversión y el mismo (o muy similar) retorno, debería tenerse en cuenta; o si, por ejemplo, la vigilancia masiva por cámaras con sistemas de reconocimiento facial puede evitarse mediante la creación de políticas legislativas de prevención general, también deberá contemplarse de forma prioritaria.

Y es que no debe olvidarse en ningún momento el peligro que supone la cesión y el tratamiento automatizado de millones de datos personales de los ciudadanos a una organización, tanto pública como privada, ya que en caso de que estos acaben en manos de alguien con intenciones ilegítimas las consecuencias podrían ser fatales e irreversibles, ya que además de la vulneración del derecho a la privacidad y protección de datos de carácter personal (artículos 7 y 8 de la Carta de Derechos Funsamentales de la UE y artículo 18.1 de la Constitucion Española), un mal uso de tales datos podría conllevar la violación de muchos otros derechos fundamentales, tal y como ya se ha advertido con anterioridad, por lo que debe prestarse especial atención y tener extrema cautela en la toma de decisiones en tal sentido.

Así, no debe olvidarse que a diferencia de lo que ocurriría con una contraseña, la numeración de una tarjeta bancaria o cualquier otro código alfanumérico, por ejemplo, que serían susceptibles de posterior cambio para evitar males mayores en caso de haber sido utilizados y tratados de forma ilegítima, los datos biométricos no pueden ser modificados, por lo que en todo caso su mal uso conduciría a situaciones de daño irreparable.

En relación con todo lo anteriormente expuesto, especialmente interesante es hacer referencia a la controvertida aplicación estadounidense ClearView, creada por Hoan Ton-That y empleada por distintas fuerzas policiales estadounidenses (incluido el FBI)[721], que está bajo la lupa de las organizaciones *pro* derechos humanos por sus posibles implicaciones en materia de privacidad.

Y es que la principal función de tal aplicación, que cuenta con una base de datos de más de tres mil millones de imágenes, es la de relacionar fotografías de personas, sin identificar, con todas aquellas imágenes suyas que se hallan colgadas en Internet, ya sea en redes sociales, páginas web, etc, con el fin de proceder a identificarlas, teniendo unos ratios de éxito muy elevados.

Así, por ejemplo, si un policía tiene la imagen de un sospechoso de comisión de un delito, del que desconoce su identidad, puede introducirla en la *app* y obtener así, en cuestión de segundos, todas las fotografías públicas de dicha persona que aparecen *on line*, junto con los correspondientes enlaces para acceder a las mismas y poder así, proceder a identificarlas (y no solo eso, ya que la información que puede obtenerse de los portales donde se hallan tales imágenes suele ir mucho más allá de la mera identificación), conllevando la elaboración de perfiles.

Ello, a pesar resultar extremadamente útil, está generando multitud de problemáticas y controversias, especialmente legales. Así, mientras el Consejero Delegado de la compañía ClearView, Hoan Ton-That, mantiene que no lleva a cabo práctica alguna que vulnere el derecho a la privacidad y a la protección de datos de los ciudadanos, siendo que únicamente recopila fotografías públicamente disponibles en Internet, de forma abierta, a las que se puede acceder desde cualquier ordenador en cualquier parte del mundo[722], numerosas organizaciones *pro* derechos humanos alegan vulneraciones del derecho a la privacidad y a la protección de los datos biométricos (entre otras la organización ACLU, de Illinois, que en marzo de 2021, junto con The Chicago Alliance Against Sexual Exploitation, The Sex Workers Outreach Project, The Illinois State Public Interest Research Group, y Mujeres Latinas en Acción, presentó una demanda a la mencionada compañía por vulnerar el derecho a la privacidad de

[721] Véase Clearview, s.f..
[722] Véase CNN, s.f..

los ciudadanos del estado de Illinois reconocido en la Ley de Privacidad de la Información Biométrica de Illinois[723], entendiendo que se habían almacenado millones de fotografías sin que sus propietarios hubieran dado su consentimiento o permiso.

En mi opinión, en la misma línea de lo establecido por Comité Europeo de Protección de Datos (CEPD), actualmente el uso de una aplicación como ClearView por parte de las autoridades para fines prevención, detección, investigación y enjuiciamiento de infracciones penales o de ejecución de sanciones penales, incluidas la protección y prevención frente a las amenazas contra la seguridad pública, no siempre resultaría conforme con el régimen de protección de datos de la UE.

Y es que, bajo mi punto de vista, la duda principal radica en si podríamos considerar que el tratamiento de las categorías especiales de datos personales llevado a cabo por dicha aplicación quedaría amparado por lo dispuesto en el artículo 13 de la LO 7/21, de 26 de mayo. En relación con ello, por un lado, entiendo que las autoridades que fueran a hacer uso de tal herramienta deberían determinar si el antedicho tratamiento resulta estrictamente necesario y, en ese caso, deberían establecer las garantías adecuadas para proteger los derechos y libertades de los interesados (lo cual, no obstante, puede dar lugar a diversas interpretaciones), siempre y cuando se cumpliera alguna de las siguientes circunstancias: previsión legal (nacional o europea), necesidad de proteger los intereses vitales, así como los derechos y libertades fundamentales de los interesados o de otras personas físicas; empleo de datos biométricos dirigidos a identificar de manera unívoca a una persona física con los fines de prevención, investigación, detección de infracciones penales, incluidas la protección y la prevención frente a las amenazas contra la seguridad pública; o empleo de datos que el interesado hubiera hecho manifiestamente públicos.

Esta última circunstancia, no obstante, considero que debería ser interpretada de forma absolutamente restrictiva, en aras de evitar la vulneración de derechos fundamentales, y en tal sentido, entiendo que el mero hecho de que una persona cuelgue una fotografía en una red social, en su propio perfil cerrado, por ejemplo, no podría entenderse como una publicación

[723] Véase Amercial Civil Liberties Union, 2020.

manifiesta (a diferencia, no obstante, de lo que podría ocurrir si la hubiera colgado en una red social con un perfil abierto al público).

Y dicho lo expuesto, pues, desde luego, lo que queda claro es que el uso de ClearView en la UE del mismo modo en que se lleva a cabo por los cuerpos policiales de EEUU, no resultaría legítimo, puesto que está libre de todo control.

En el ámbito del reconocimiento de emociones, es de especial relevancia hacer referencia a lo que ocurre en China, en concreto en la región de Xinjiang (Sinkiang). Y es que, según un testigo anónimo, cuyo relato fue comprobado por la ONG Human Rights Watch, el gobierno chino está probando, sin consentimiento, sistemas de reconocimiento de emociones entre la población uigur.

En relación con ello, dicho testigo aseguró al portal de noticias BBC que él, junto a otros profesionales, colocó cámaras de detección de emociones a tres metros de los distintos sujetos, aseverando "*es similar a un detector de mentiras pero con tecnología mucho más avanzada*". Y es que según lo manifestado por este testigo, tales sistemas de reconocimiento de emociones se emplean fundamentalmente en los interrogatorios policiales de los uigures que resultan detenidos para verificar si mienten, estando el sistema de IA entrenado para analizar y detectar cualquier mínimo cambio en la expresión facial (inclusive en los poros de la piel). Además, el mencionado testigo explicó que los agentes emplean "sillas de sujeción" que bloquean las muñecas y los tobillos de estos con elementos de metal.[724]

Y ello, desde luego, resulta absolutamente inaceptable. Sin duda, desde la perspectiva de la protección de datos personales, puesto que su tratamiento no resulta ni consentido ni legalmente justificado, y también desde la perspectiva de la ausencia absoluta de transparencia del proceso (de hecho, el gobierno chino lo niega) y, por ende, de la imposibilidad de comprobar la calidad real del sistema y su grado de seguridad y precisión, constituyendo bajo mi punto de vista una vulneración de derechos humanos en toda regla.

c) Posibles brechas de seguridad

[724] Véase Wakefield, 2021.

En cuarto lugar, es importante hacer referencia a los riesgos que los analizados sistemas de reconocimiento facial tienen en materia de seguridad.

Así, resulta imperante la necesidad de que tales sistemas cuenten con medidas eficaces y seguras para que las bases de datos, que albergan información biométrica (y, por ende, sensible), resulten en todo momento debidamente protegidas, habida cuenta de laenorme exposición que tienen a filtraciones y ciberataques, como el que tuvo lugar, por ejemplo, en 2019, en la base de datos del Departamento de Aduanas y Fronteras de EEUU, que dejó al descubierto varios miles de imágenes de viajeros sin autorización previa.[725]

Y es que es fundamental que los sistemas de IA que emplean datos biométricos cuenten con mecanismos de protección y prevención de las potenciales consecuencias de su mala utilización, ya que tal y como ya se ha ido afirmando a lo largo de la presente tesis doctoral, cuando todo va según lo previsto, la tecnología puede ser una bendición, pero cuando las circunstancias cambian y se tuercen, esta puede convertirse en un arma extremadamente peligrosa.

Así, la seguridad de los sistemas es clave para el éxito de los mismos y, sobre todo, para su uso de forma garantista y confiable en el ámbito de la investigación criminal, especialmente sensible, ya que el riesgo que implica que la información contenida en los sistemas sea empleada para otras finalidades distintas de las legalmente previstas, en el caso de que se produzca un fallo de seguridad, es altamente preocupante y debe quedar reducido al mínimo.

Por ejemplo, imaginemos las terribles consecuencias que tendría el hecho de que un grupo de *hackers* llevara a cabo un ciberataque que consiguiera dejar al descubierto las imágenes y los datos biométricos de miles de personas sospechosas o investigadas por la comisión de algún delito que constaran almacenadas en la base de datos de un sistema de reconocimiento facial y, con el ánimo de sembrar caos, las publicara bajo el título de "delincuentes peligrosos" en un portal de Internet al que tuvieran acceso millones de usuarios. ¿Cómo se gestionaría la reparación del enorme daño al honor causado a esas miles de personas amparadas por el

[725] Véase Kanno-Youngs & Sanger, 2019.

principio de presunción de inocencia que verían asociados sus rostros con tal grave calificación? Ciertamente, y a pesar de la existencia de herramientas legales para castigar a los autores de los mencionados hechos, el perjuicio sería incalculable (especialmente en el caso de aquellas personas cuyos rostros hubieran sido introducidos por error en tal base de datos).

O bien, imaginemos que un grupo terrorista, con el ánimo de conocer el estado de la investigación de unos hechos delictivos cometidos por sus miembros, decide atacar el sistema de reconocimiento facial empleado por la división antiterrorista de la policía y descubrir, así, qué personas se hallan etiquetadas como sospechosas o investigadas por comisión de delitos terroristas para, en caso de verificar que figuran en la misma, destruir la información biométrica o adoptar decisiones tales como la de huir del país tratando de burlar la precisión de los sistemas de reconocimiento facial (con gorros, gafas de sol, bufandas, etc) antes de ser detectados por estos.

Y estos solo son algunos de los millones de terribles y nocivos escenarios que podrían darse en caso de que los *software* de reconocimiento facial no estén dotados de unos sistemas de seguridad altamente infalibles.

En relación con lo expuesto, procede hacer referencia a las obligaciones que, en materia de seguridad, se contienen en la LO 7/21, de 26 de mayo, en concreto en la Sección 2ª del Capítulo IV (artículos 37 a 39).

Así, el artículo 37 dispone una obligación para el responsable y el encargado del tratamiento de los datos biométricos, que debe tener en cuenta el estado de la técnica y los costes de aplicación, y la naturaleza, el alcance, el contexto y los fines del tratamiento, así como los niveles de riesgo para los derechos y libertades de las personas físicas para aplicar: la aplicación de "*medidas técnicas y organizativas apropiadas para garantizar un nivel de seguridad adecuado, especialmente en lo relativo al tratamiento de las categorías de datos personales a las que se refiere el artículo 13. En particular, deberán aplicar a los tratamientos de datos personales las medidas incluidas en el Esquema Nacional de Seguridad.*"

Y, en concreto, respecto del tratamiento automatizado de datos biométricos, el responsable o encargado del tratamiento, a raíz de una evaluación de los riesgos, "*pondrá en práctica medidas de control con el siguiente propósito:*

a) En el control de acceso a los equipamientos, denegar el acceso a personas no autorizadas a los equipamientos utilizados para el tratamiento.
b) En el control de los soportes de datos, impedir que estos puedan ser leídos, copiados, modificados o cancelados por personas no autorizadas.
c) En el control del almacenamiento, impedir que se introduzcan sin autorización datos personales, o que estos puedan inspeccionarse, modificarse o suprimirse sin autorización.
d) En el control de los usuarios, impedir que los sistemas de tratamiento automatizado puedan ser utilizados por personas no autorizadas por medio de instalaciones de transmisión de datos.
e) En el control del acceso a los datos, garantizar que las personas autorizadas a utilizar un sistema de tratamiento automatizado, sólo puedan tener acceso a los datos personales para los que han sido autorizados.
f) En el control de la transmisión, garantizar que sea posible verificar y establecer a qué organismos se han transmitido o pueden transmitirse, o a cuya disposición pueden ponerse los datos personales mediante equipamientos de comunicación de datos.
g) En el control de la introducción, garantizar que pueda verificarse y constatarse, a posteriori, qué datos personales se han introducido en los sistemas de tratamiento automatizado, en qué momento y quién los ha introducido.
h) En el control del transporte, impedir que durante las transferencias de datos personales o durante el transporte de soportes de datos, los datos personales puedan ser leídos, copiados, modificados o suprimidos sin autorización.
i) En el control de restablecimiento, garantizar que los sistemas instalados puedan restablecerse en caso de interrupción.
j) En el control de fiabilidad e integridad, garantizar que las funciones del sistema no presenten defectos, que los errores de funcionamiento sean señalados y que los datos personales almacenados no se degraden por fallos de funcionamiento del sistema."

Por su parte, los artículos 38 y 39, respectivamente, establecen la obligación de notificación a la autoridad de protección de datos y al interesado en casos de violación de la seguridad de los datos personales.

Así, el primero de tales preceptos dispone que: "*1. Cualquier violación de la seguridad de los datos personales será notificada por el responsable del tratamiento a la autoridad de protección de datos competente, a menos que sea improbable que la violación de la seguridad de los datos personales constituya un peligro para los derechos y las libertades de las personas físicas.*

La notificación deberá realizarse en el plazo de las setenta y dos horas siguientes al momento en que se haya tenido constancia de ella. En caso contrario, deberá ir acompañada de los motivos de la dilación.

2. El encargado del tratamiento notificará, sin dilación indebida, al responsable del tratamiento, las violaciones de la seguridad de los datos personales de las que tenga conocimiento.

3. La notificación contemplada en el apartado 1 deberá, al menos:

a) Referir la naturaleza de la violación de la seguridad de los datos personales, incluyendo, cuando sea posible, las categorías y el número aproximado de personas afectadas, así como las categorías y el número aproximado de registros de datos personales afectados por la violación de la seguridad.

b) Comunicar el nombre y los datos de contacto del delegado de protección de datos o de otro punto de contacto en el que pueda obtenerse más información.

c) Detallar las posibles consecuencias de la violación de la seguridad de los datos personales.

d) Describir las medidas adoptadas o propuestas por el responsable del tratamiento para poner remedio a la violación de la seguridad de los datos personales, incluyendo, si procede, las medidas adoptadas para mitigar sus posibles efectos negativos.

4. Si no fuera posible facilitar la información simultáneamente, se podrá facilitar de forma progresiva, a medida que se disponga de ella.

5. El responsable del tratamiento documentará cualquier violación de la seguridad de los datos personales, incluidos los hechos relativos a dicha violación, sus efectos y las medidas correctivas adoptadas.

Dicha documentación estará a disposición de la autoridad de protección de datos competente al objeto de verificar el cumplimiento de lo dispuesto en este artículo.

6. Cuando la violación de la seguridad de los datos personales afecte a datos que hayan sido transmitidos por el responsable del tratamiento o al responsable del tratamiento de otro Estado miembro de la Unión Europea, la información recogida en el apartado 3 se comunicará al responsable del tratamiento de dicho Estado.

7. Todas las actividades relacionadas en este artículo se realizarán sin dilaciones indebidas."

Y el segundo de tales preceptos establece: "*1. Cuando existan indicios de que una violación de la seguridad de los datos personales supondría un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento comunicará al interesado, sin dilación indebida, la violación de la seguridad de los datos personales.*

2. La comunicación al interesado describirá con lenguaje claro, sencillo y accesible conforme a sus circunstancias y capacidades, la naturaleza de la violación de la seguridad de los datos personales y contendrá, al menos, la información y las medidas a las que se refiere el artículo 38.3. b), c) y d).

3. No se efectuará la comunicación al interesado que prevé el apartado 1 cuando se cumpla alguna de las condiciones siguientes:

a) Que el responsable del tratamiento haya adoptado medidas apropiadas de protección técnica y organizativa y dichas medidas se hayan aplicado a los datos personales afectados por la violación de la seguridad antes de la misma, en particular, aquellas que hagan ininteligibles los datos personales para cualquier persona que no esté autorizada a acceder a ellos, como en el caso del cifrado.

b) Que el responsable del tratamiento haya tomado medidas ulteriores para garantizar que no se materialice el alto nivel de riesgo para los derechos y libertades del interesado a que hace referencia el apartado 1.

c) Que suponga un esfuerzo desproporcionado, en cuyo caso, se optará por su publicación en el boletín oficial correspondiente, en la sede electrónica del responsable del tratamiento o en otro canal oficial que permita una comunicación efectiva con el interesado.

4. En el supuesto de que el responsable del tratamiento no haya comunicado al interesado la violación de la seguridad de los datos personales, la autoridad de protección de datos competente, una vez valorada la existencia de un alto nivel de riesgo, podrá exigirle que proceda a dicha comunicación, o bien que determine la concurrencia de alguna de las condiciones previstas en el apartado 3.

5. La comunicación al interesado referida en el apartado 1 podrá aplazarse, limitarse u omitirse con sujeción a las condiciones y por los motivos previstos en el artículo 24."

Por su parte, en la Propuesta de Reglamento de la IA publicada por la Comisión Europea el 21 de abril de 2021 se introducen asimismo obligaciones en materia de seguridad, especialmente en relación con los sistemas de IA de identificación biométrica remota en tiempo real, calificados de "alto riesgo", tal y como ya se ha advertido en secciones anteriores.

Y es que el artículo 15, por ejemplo, establece que:

"*Los sistemas de Inteligencia Artificial de alto riesgo se diseñarán y desarrollarán de manera que alcancen, a la luz de su finalidad prevista, un nivel adecuado de precisión, robustez y ciberseguridad, y funcionarán de manera consistente en tales aspectos a lo largo de su ciclo de vida*. (…)

3. Los sistemas de Inteligencia Artificial de alto riesgo deberán ser resistentes a los errores, fallos o incoherencias que puedan producirse en el sistema o en el entorno en el que este funciona, especialmente como consecuencia de su interacción con personas físicas u otros sistemas.

La solidez de los sistemas de IA de alto riesgo se puede lograr mediante soluciones de redundancia técnica, que pueden incluir planes de respaldo o a prueba de fallos. (…)

Los sistemas de IA de alto riesgo serán resistentes a los intentos de terceros no autorizados de alterar su uso o rendimiento mediante la explotación de las vulnerabilidades del sistema.

Las soluciones técnicas destinadas a garantizar la ciberseguridad de los sistemas de IA de alto riesgo deberán ser adecuadas a las circunstancias y riesgos relevantes.

Las soluciones técnicas para abordar las vulnerabilidades específicas de la IA deben incluir, cuando corresponda, medidas para prevenir y controlar los ataques que intenten manipular el conjunto de datos de entrenamiento ("envenenamiento de datos"), los efectos diseñados para hacer que el modelo cometa errores ("ejemplos contradictorios") o los defectos del modelo."

d) Posible falta de transparencia

Y, finalmente, pero no por ello menos importante, es fundamental hacer referencia a los posibles riesgos jurídicos que giran en torno a nivel de transparencia de los sistemas de IA de reconocimiento biométrico.

Y es que, sin lugar a dudas, únicamente la transparencia de los mencionados sistemas puede permitir la detección de los potenciales peligros y riesgos que estas entrañan realmente en cada caso concreto, erigiéndose así como su cualidad más importante, habida cuenta de que su concurrencia es *conditio sine qua non* para llevar a cabo su correspondiente examen. Así, de no existir transparencia respecto del modo de funcionar de tales sistemas y de los datos que estos emplean, difícilmente podrá advertirse por los humanos cuál es, en su caso, la causa de su falta de precisión; si existen o no sesgos; si se da o no una vulneración del derecho a la privacidad y a la protección de datos; o si concurre una brecha de seguridad que proceda salvar.

En tal sentido, por ejemplo, la organización estadounidense ACLU, en su división de Michigan (EEUU), ante la sospecha de que la policía de dicho estado empleaba sistemas de reconocimiento facial que incluían sesgos raciales y, en ocasiones, arrojaban resultados erróneos, emitió una solicitud al Departamento de Interior y a la Policía Estatal para que, de forma transparente, publicaran datos al respecto y, en concreto: aportaran información sobre

la base de datos "*Statewide Network of Agency Photos*" (SNAP), haciendo especial referencia a su contenido y a la justificación legal existente para que el Departamento de Interior compartiera automáticamente con la Policía Estatal las fotografías del carné de conducir de los ciudadanos para incluirlas en tal base de datos; información sobre qué búsquedas de reconocimiento facial realizaba la Policía Estatal en dicha base de datos, así como los acuerdos de la Policía estatal (MSP) para compartir el acceso a la mencionada base de datos con las agencias policiales locales; información sobre si se compartían recursos o datos de reconocimiento facial con las autoridades federales de inmigración y, en su caso, cómo ello se llevaba a cabo; información sobre la formación que recibían los oficiales de policía encargados de manejar tales sistemas; e información sobre cómo se protegía y se auditaba tal información.

En el ámbito del proceso penal, tal y como se ha advertido ya en secciones anteriores, la necesidad de transparencia cobra especial relevancia, habida cuenta de que la falta de la misma en un sistema de IA de reconocimiento de datos biométricos puede suponer un choque frontal con el derecho fundamental a la defensa y al proceso debido, con todas las garantías. Así, en aquellos procedimientos penales en que se hayan empleado tal clase de herramientas tecnológicas para incriminar a una o varias personas, tanto en fase de instrucción como en fase de plenario, estas (así como la acusación particular, en su caso, y el Ministerio Fiscal) deben tener acceso pleno a toda la información técnica, de contenido y de funcionamiento de las mismas, en aras de poder articular su defensa y ejercer su derecho de contradicción, puesto que, de lo contrario, verían limitados sus derechos constitucionalmente reconocidos.

Ello, no obstante, por desgracia, no resulta siempre tan obvio. Y es que, por ejemplo, una corte de apelaciones de Florida (EEUU) manifestó su postura contraria en el conocido caso Lynch.[726]

Así, en 2015, en Jacksonville (Florida) dos policías encubiertos compraron cocaína en la vía pública a un hombre negro y, en lugar de proceder a su detención en el acto, usaron un teléfono viejo para tomar varias instantáneas de él mientras fingían estar en una llamada, por lo que la calidad de las imágenes brillaba por su ausencia. No obstante, tales fotografías fueron

[726] Véase Justia Law, 2018.

empleadas por un analista criminal, que usó un sistema de IA de reconocimiento facial (denominado FACES) para verificar su identidad. Tal herramienta arrojó varios resultados, coincidencias o *match*, correspondiendo el primero de ellos (con tan solo una estrella de confianza[727]) al ciudadano de raza negra Willie Allen Lynch Lynch, quien enseguida fue arrestado y acusado de la venta de drogas.

No obstante, con carácter previo al juicio, cuando el Letrado de la defensa, que mantenía que existía una equivocación y que su cliente era inocente, interrogó al analista criminal que empleó el sistema de IA en el que se basaba la acusación (puesto que los agentes de policía no habían sido capaces de reconocer al investigado previamente, aunque en el acto del plenario aseguraron que era él el proveedor de la droga, sugestionados por los resultados del mencionado sistema), este reconoció que no entendía cómo funcionaba el algoritmo de reconocimiento facial utilizado, por lo que la defensa de Lynch argumentó que era necesario poder probar la confiabilidad y la precisión de tal herramienta. Así, dicho Letrado solicitó ver las otras imágenes que el programa de IA detectó como posibles coincidencias, si bien tanto la Fiscalía como el Tribunal se negaron, alegando sin lógica alguna, en mi opinión, que los agentes de policía tampoco las habían visto y, por ende, ello carecía de sentido.

Como consecuencia de ello, las organizaciones ACLU y Electronic Frontier Foundation, así como el Georgetown Center on Privacy and Technology y la ONG The Innocence Project presentaron un escrito instando al Tribunal Supremo de Florida a pronunciarse sobre la necesidad de transparencia máxima en casos de utilización de sistemas de identificación biométrica para identificar a presuntos delincuentes y proceder a su condena.

En relación con ello, la LO 7/21, de 26 de mayo, establece obligaciones dirigidas a potenciar y garantizar al máximo la transparencia de los sistemas empleados por las autoridades para fines de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales.

Así, en el artículo 20 se dispone que: "*1. El responsable del tratamiento deberá facilitar al interesado, de forma concisa, inteligible, de fácil acceso y con lenguaje claro y sencillo para*

[727] El sistema calificaba la probabilidad de éxito de una coincidencia asignando estrellas a las imágenes resultantes.

todas las personas, incluidas aquellas con discapacidad, toda la información contemplada en el artículo 21, así como la derivada de los artículos 14, 22 a 26 y 39. (…)."

Y es que, con interés, asimismo, en materia de transparencia, el artículo 21, determina que: "1. *El responsable del tratamiento de los datos pondrá a disposición del interesado, al menos, la siguiente información:*

a) La identificación del responsable del tratamiento y sus datos de contacto.
b) Los datos de contacto del delegado de protección de datos, en su caso.
c) Los fines del tratamiento a los que se destinen los datos personales.
d) El derecho a presentar una reclamación ante la autoridad de protección de datos competente y los datos de contacto de la misma.
e) El derecho a solicitar del responsable del tratamiento el acceso a los datos personales relativos al interesado y su rectificación, supresión o la limitación de su tratamiento.

2. Además de la información a la que se refiere el apartado 1, atendiendo a las circunstancias del caso concreto, el responsable del tratamiento proporcionará al interesado la siguiente información adicional para permitir el ejercicio de sus derechos:

a) La base jurídica del tratamiento.
b) El plazo durante el cual se conservarán los datos personales o, cuando esto no sea posible, los criterios utilizados para determinar ese plazo.
c) Las categorías de destinatarios de los datos personales, cuando corresponda, en particular, los establecidos en Estados que no sean miembros de la Unión Europea u organizaciones internacionales.
d) Cualquier otra información necesaria, en especial, cuando los datos personales se hayan recogido sin conocimiento del interesado."

Y el artículo 22 dispone en su apartado 1 que: "*1. El interesado tendrá derecho a obtener del responsable del tratamiento confirmación de si se están tratando o no datos personales que le conciernen. En caso de que se confirme el tratamiento, el interesado tendrá derecho a acceder a dichos datos personales, así como a la siguiente información:*

a) Los fines y la base jurídica del tratamiento.
b) Las categorías de datos personales de que se trate.
c) Los destinatarios o las categorías de destinatarios a quienes hayan sido comunicados los datos personales, en particular, los destinatarios establecidos en Estados que no sean miembros de la Unión Europea u organizaciones internacionales.
d) El plazo de conservación de los datos personales, cuando sea posible, o, en caso contrario, los criterios utilizados para determinar dicho plazo.
e) La existencia del derecho a solicitar del responsable del tratamiento la rectificación o supresión de los datos personales relativos al interesado o la limitación de su tratamiento.
f) El derecho a presentar una reclamación ante la autoridad de protección de datos competente y los datos de contacto de la misma.
g) La comunicación de los datos personales objeto de tratamiento, así como cualquier información disponible sobre su origen, sin revelar la identidad de ninguna persona física, en especial en el caso de fuentes confidenciales."

En el ámbito de la UE, respecto del uso de los sistemas de IA de identificación biométrica remota en tiempo real (distintos al del caso anterior), en aquellos casos en que este resulta permitido, la ya mencionada Propuesta de regulación de la IA publicada por la Comisión Europea el 21 de abril de 2021, establece obligaciones específicas.

Así, por un lado, el artículo 11 establece la obligación de que tales sistemas, con carácter previo a ser comercializados o puestos en funcionamiento, cuenten con documentación técnica que proporcione información sobre el cumplimiento de los requisitos exigidos por la normativa, en los términos previstos en el Anexo IV; por otro lado, el artículo 12 dispone la obligación de que dichos sistemas se diseñen y se desarrollen con capacidades que permitan el registro automático de eventos ("registros") durante su funcionamiento, garantizando así un nivel de trazabilidad, a lo largo de su ciclo de vida, que sea apropiado para el propósito preestablecido del sistema; y, finalmente, el artículo 13 establece una obligación específica de transparencia y provisión de información a los usuarios, y en concreto determina: "*1. Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de tal forma que garanticen que su funcionamiento sea lo suficientemente transparente para permitir a los usuarios interpretar la salida del sistema y utilizarla de forma adecuada. Se garantizarán un tipo y un*

grado de transparencia adecuados con el fin de lograr el cumplimiento de las obligaciones correspondientes al usuario y al proveedor establecidas en el Capítulo 3 del presente Título.

2. Los sistemas de IA de alto riesgo deberán ir acompañados de instrucciones de uso en un formato digital apropiado, o de otro modo que incluyan información concisa, completa, correcta y clara que sea relevante, accesible y comprensible para los usuarios."

B- Herramientas de IA que emplean técnicas de Procesamiento del Lenguaje Natural (PLN)

B.1. *Concepto*

Las técnicas de PLN tienen por objeto traducir el lenguaje humano (hablado o escrito) en un lenguaje que "la máquina" (o, de forma más específica, el algoritmo) pueda entender. Así, en aquellos casos en que el interlocutor remita información oral, será necesario que la máquina transforme la voz en palabras y, posteriormente, interprete su significado; y, en aquellos casos en que el interlocutor remita información escrita, únicamente será necesario que la máquina interprete el significado de las palabras escritas.

Esta tecnología, que sin duda tiende a imitar, una vez más, funciones cognitivas del cerebro humano, es empleada por diversas herramientas de IA, algunas de las cuales pueden resultar de enorme utilidad en el ámbito de la investigación criminal.

A pesar de la enorme proliferación que están teniendo este tipo de sistemas durante los últimos años, principalmente gracias a la evolución y a la madurez de las técnicas de aprendizaje automático, lo cierto es que ya empezaron a emplearse en los años 70. No obstante, no fue hasta 1996 cuando IBM presentó el primer *software* de reconocimiento del habla, el denominado "*VoiceType Simply Speaking*", que permitía dictar de forma oral textos al ordenador a una velocidad de setenta a cien palabras por minuto, con posibilidad de corrección directa.[728]

[728] Véase Dealer World, 1996.

Desde entonces, no obstante, la calidad y la precisión de los mencionados sistemas ha mejorado a pasos agigantados, si bien todavía incurren en errores que deben eliminarse mediante la creación de bases de datos masivas, de calidad, el diseño de algoritmos sofisticados, precisos y eficientes, el uso de memorias RAM ("*Random Acces Memory*") más potentes, y la introducción de procesadores más veloces que los actuales.

Y es que tales sistemas de IA lo que hacen es convertir una señal de audio analógico en una señal digital que deberá ser descifrada por el dispositvo de que se trate mediante la extracción, reconocimiento y comparación de patrones con los de la información (a saber, sílabas, palabras, expresiones etc) contenida en una base de datos digital. Así, cuanto más extensa sea la capacidad de memoria del dispositivo y, por ende, la mencionada base de datos, menores serán los errores de precisión del sistema.

B.2. *Subclases*

a) Chatbots

a.1) *Concepto*

A modo conceptual, un *chatbot* es una aplicación de *software* que tiene por objeto mantener una conversación *on line* de voz a texto, simplemente de texto, o con combinación de ambas (dependiendo de los distintos canales de entrada y salida de información), mediante técnicas de PLN. Y es que tales herramientas, generalmente están orientados a mantener conversaciones y emitir respuestas preconstituidas ante un interlocutor humano, pudiendo sus usos ser múltiples y muy variados.

Dentro de tal clase de herramientas (potencialmente útiles para la investigación de las causas), debemos centrarnos en los denominados *chatbots* cognitivos o *Smart Chatbots*, que emplean IA (en concreto, *Machine Learning*) para comprender y procesar el lenguaje natural mediante las mencionadas técnicas de PLN.

Tales *chatbots* cognitivos cuentan con la capacidad de entender el contexto de las conversaciones, interpretar la intención del usuario y elaborar respuestas con sentido,

otorgando a la interacción una naturalidad y espontaneidad muy elevadas. Además, al emplear técnicas de *Machine Learning*, los mencionados *chatbots* van aprendiendo por sí mismos conforme van siendo utilizados, por lo que a mayor frecuencia de interacción, mejor calidad de las comunicaciones.

Es interesante poner de manifiesto, también, que los *Smart Chatbots* tienen capacidad para aprender a interactuar con usuarios concretos, pudiendo incluso llegar a ser proactivos, ofreciéndoles respuestas y propuestas personalizadas con base en sus patrones de comportamiento. Además, pueden llevar a cabo operaciones concretas de manera automática, como por ejemplo, llamar a la policía.

Igualmente, es relevante advertir que este tipo de herramientas pueden incluir, para aumentar su utilidad, tecnología de reconocimiento de voz, a los efectos de que el *software*, en aquellos casos en que recibe mensajes orales, pueda detectar quién es su interlocutor.

Hoy en día el uso de tales sistemas es principalmente comercial, ya que este tipo de aplicaciones son empleadas por multitud de empresas de todo tipo, especialmente en las áreas de atención al cliente, tales como BBVA Argentina, que incorporó a su *chatbot* Lucía, o ŠKODA, filial del Grupo Volkswagen, que incorporó a su *chatbot* Laura; y, especialmente por empresas tecnológicas, como servicio de asistencia virtual, y es que, por ejemplo, el ya mencionado Siri, de los teléfonos Apple, o Google Assistant, de los teléfonos Android, pueden ser también considerados *chatbots*.

No obstante, en la esfera de la investigación de delitos, a pesar de su enorme potencial, actualmente su uso es prácticamente nulo, aunque se están empezando a hacer avances.

Así, especial mención merece el *chatbot* denominado "*CopsBot*", presentado en el "*Smart India Hackathon 2020*", organizado por el entonces Ministerio de Desarrollo de Recursos Humanos (actual Ministerio de Educación). Dicha herramienta, muy novedosa, tiene como objetivo facilitar a los ciudadanos la interposición de denuncias y permitirles tener conocimiento de las potenciales amenazas delictivas existentes en su zona. Y es que mediante la tecnología de IA incorporada no solo permite a los usuarios remitir a la policía, de forma

directa, información para registrar una denuncia, sino que además les posibilita el envío de su ubicación al centro de comando en caso de emergencia y la remisión de mensajes de voz.

Y, sin duda, es también merecedora de atención la nueva aplicación de la asistente virtual Alexa, de Amazon, denominada "*Glow Up, Damas*", creada por la brasileña Danielle Vantini, ganadora del "*Global Amazon Alexa Skills Challenge 2021*"[729], que tiene como objetivo reducir los casos de violencia de género.[730]

La mencionada herramienta que, no obstante, todavía se halla en fase de pruebas, opera con IA y está configurada para ser activada por comandos y respuestas codificadas y, en caso de percibir amenaza, llamar a un taxi autónomo gratuito, proporcionando así una vía de escape rápida para las víctimas.

a.2) ***Posibles utilidades en la instrucción de las causas***

a.2.1) Auxilio para víctimas con medidas y penas de prohibición de aproximación a su favor

En la actualidad, en caso de que una víctima que cuente con una orden de alejamiento a su favor (como medida cautelar o como condena firme) necesite pedir ayuda de forma inmediata, lo único que puede hacer es, o bien gritar para ser oída por la gente de su alrededor, o bien escribir o llamar a algun familiar o amigo, o bien llamar a los servicios de emergencias o a la policía.

En los casos de las mujeres víctimas de violencia de género, además, en ocasiones los jueces acuerdan la imposición de un dispositivo de control telemático (conocidos informalmente como "pulseras") que, mediante el uso de la tecnología GPS y la radiofrencuencia proporciona información a tiempo real sobre el cumplimiento de las medidas de alejamiento y sus incidencias, siendo que, entre otros, en caso de detectar que el investigado o condenado se está aproximando a la víctima incumpliendo así la prohibición impuesta, o en caso de advertir que el brazalete ha sido roto o manipulado, genera un aviso a un centro de control (el denominado Centro de Control COMETA) que efectúa las comunicaciones detalladas en el

[729] Un concurso organizado por Amazon de forma anual.
[730] Véase Shimabukuro, 2021.

Protocolo de actuación para cada tipo de alerta y, además, remite un informe a las autoridades policiales responsables de la protección de la víctima, al Ministerio Fiscal y a la Autoridad Judicial.[731]

No obstante, la realidad es que tal sistema se emplea en contadas ocasiones, ya que genera múltiples incidencias que en algunos casos producen más estrés que ayuda a las víctimas, causando justamente el efecto contrario de lo que se pretende conseguir.

Así, y sin perjuicio de que puedan seguir utilizándose los antedichos métodos de protección (los tradicionales y el del control telemático), lo cierto es que los *chatbots*, bajo mi punto de vista, podrían ofrecer un eficaz auxilio extra a las víctimas. Y es que en mi opinión, la instalación en sus teléfonos móviles o en otros dispositivos (que podrían incluso ser los brazaletes de control telemático) de una herramienta de IA capaz de comprender sus mensajes (e, incluso, reconocer su voz y la de su agresor) y ejecutar órdenes de forma inmediata, sería más que útil tanto para la protección de tales víctimas como para el éxito de la instrucción de las causas.

Así, por un lado, pensemos en una víctima de violencia de género que se halla paseando por una zona solitaria y ve de lejos a su ex pareja, sobre la que pesa una condena firme de prohibición de aproximación a menos de mil metros. Imaginemos, asimismo, que en ese momento se pone tan nerviosa que no es capaz de marcar el teléfono de emergencias, por lo que procede a pronunciar la palabra "policía" (o cualquier palabra clave, neutra, que no permita al potencial agresor conocer que la víctima ha requerido a las autoridades policiales) y el dispositivo reconoce su voz, la geolocaliza y directamente se pone en contacto con la policía para que acuda al lugar de los hechos lo más rápido posible.

O imaginemos que, para no resultar oída por su potencial agresor, simplemente abre el *chatbot*, presiona la tecla "P" y este de forma inmediata comprende el mensaje y ejecuta la orden de geolocalizarla y llamar a la policía.

En tales casos, desde luego, de no haberse empleado tal sistema de IA, y si la víctima no hubiera contado con un brazalete de control telemático (lo cual es muy frecuente), lo más

[731] Véase Delegación del Gobierno contra la Violencia de Género, s.f..

seguro es que el aviso a la policía no se hubiera producido o se hubiera realizado demasiado tarde, con las nefastas consecuencias que ello podría conllevar.

Por otro lado, imaginemos que en el mismo caso, el condenado se aproxima a la víctima y la insulta, por lo que el sistema, al reconocer su voz, de forma automática genera una alerta a la policía remitiéndole, asimismo, la geolocalización de la víctima. Y es que en tal caso, como vemos, no haría ni falta que la víctima pronunciara palabra clave alguna, puesto que la propia herramienta ya estaría configurada para, en caso de detectar la voz del condenado, grabarla y ejecutar la orden de llamar a la policía, lo cual entiendo que podría resultar tremendamente útil.

Además, entiendo que, tal y como he avanzado, si tales sistemas no solo detectan, reconocen las voces y ejecutan órdenes sino que además las graban, el provecho para la instrucción de las causas podría ser enorme, siendo que esta quedaría reducida a la mínima expresión, habida cuenta de que al juez de instrucción ya le llegaría, de entrada, un informe en el que constarían la grabación (con fecha y hora exacta), el reconocimiento por el sistema de la voz del autor de los hechos y de la víctima, en su caso, y la geolocalización de estos.

a.2.2) Información en tiempo real a las Fuerzas y Cuerpos de Seguridad

En diversas ocasiones los ciudadanos, sin previo aviso, se convierten en testigos de la comisión de ilícitos penales, lo cual les implica de lleno en las causas judiciales que estos posteriormente generan, convirtiéndose muchas veces en piezas clave para la resolución de los casos.

En algunos supuestos, tales testigos se hallan tan inmersos en la perpetración de los delitos que tienen serias dificultades para comunicarse con la policía sin ser vistos o escuchados por los autores de los hechos, por lo que a veces ven frustrada su voluntad de parar o prevenir a tiempo la comisión de un ilícito penal, habida cuenta de que su integridad física y/o moral está en juego.

No obstante, en caso de que la policía contara con una aplicación que tuviera incluida una función de *chatbot*, todo resultaría mucho más sencillo.

Así, por un lado, imaginemos que una chica se halla en una estación de tren y observa cómo un individuo, cuando cree que nadie le ve, manipula un paquete sospechoso de ser explosivo. Pensemos que en el andén solamente están la mencionada joven y dos chicos más que se encuentran alejados e inmersos en la música de sus auriculares, por lo que el silencio es absoluto. Figurémonos, también, que dicha muchacha se pone muy nerviosa y no quiere perder de vista al sospechoso, por lo que no ve viable ir a avisar al personal de seguridad de la estación, si bien tampoco puede llamar por teléfono a la policía porque sería descubierta por el potencial terrorista. Imaginemos, no obstante, que la testigo cuenta con un aplicación en su *smartphone* que permite, a través de un *chatbot*, contactar de forma inmediata con la policía, que no solo interpreta y entiende sus mensajes sino que además, la geolocaliza. Y, supongamos, que dicha joven decide poner en conocimiento de las autoridades policiales lo que está ocurriendo a través del *chatbot*, que de forma inmediata la geolocaliza y lanza aviso a la patrulla más cercana para que acuda al lugar de los hechos de la forma más rápida posible, lo cual permite que los agentes consigan localizar al sospechoso y evitar así una catástrofe.

En tal caso, desde luego, de no haber sido por el mencionado *chatbot*, posiblemente la chica hubiera tenido que perder de vista al sospechoso para ir a avisar al personal de seguridad de la estación, con el riesgo de ser vista por este durante su desplazamiento, o hubiera tenido que avisar por mensaje o *chat* a algún familiar o amigo y, seguramente, siendo que se trataba de un tema grave y delicado, dicho personal o dicho pariente o amigo hubieran tenido que llamar a la policía, decirle dónde estaban ocurriendo los hechos y esperar hasta que se hubiera puesto el asunto en conocimiento de la patrulla más cercana, lo cual hubiera conllevado una demora considerable y, muy probablemente, excesiva para el buen fin de la intervención.

Por otro lado, imaginemos que un individuo está durmiendo en la habitación principal de su domicilio cuando escucha ruidos en la zona de la entrada y se dispone a mirar por la ventana, momento en el que se percata de que hay tres encapuchados intentando acceder a la vivienda. Pensemos, no obstante, que ello ocurre de madrugada y, por ende, reina el silencio en la zona, por lo que en caso de llamar a la policía este sería escuchado por los asaltantes. Figurémonos, no obstante, que dicho ciudadano cuenta con una aplicación en su *smartphone* que permite, a través de un *chatbot*, contactar de forma inmediata con la policía, que no solo interpreta y entiende sus mensajes sino que además, lo geolocaliza. Y, supongamos, que este individuo decide poner en conocimiento de las autoridades policiales lo que está ocurriendo a través de

tal *chatbot*, que de forma inmediata la geolocaliza y lanza aviso a la patrulla más cercana para que acuda al lugar de los hechos de la forma más rápida posible, lo cual permite que los agentes consigan detener *in fraganti* a los autores del robo con fuerza.

En tal caso, desde luego, de no haber sido por el mencionado *chatbot*, posiblemente el mencionado ciudadano hubiera tenido que correr el riesgo de llamar a la policía, lo cual podría haber comprometido su seguridad y el satisfactorio final de la intervención.

Además, en relación con lo expuesto procede advertir que las conversaciones mantenidas a través de los antedichos *chatbots* sin duda tendrían el valor de prueba preconstituida en los procedimientos judiciales que posteriormente se iniciaran, lo cual podría aportar un plus de objetividad y rigor a la instrucción de las causas.

a.3) Identificación de autores de delitos contra la libertad e indemnidad sexual de los menores y prevención de la comisión de tal clase de delitos

Y es que procede hacer referencia a la existencia de *chatbots* creados con la finalidad de ser lanzados en la red de Internet y atraer a pedófilos y otros delincuentes sexuales que pretenden pagar por llevar a cabo prácticas sexuales virtuales (generalmente a través de una *webcam*) y, en ocasiones, también físicas, con menores.

En relación con ello procede poner de manifiesto que ya en 2011 Shawn Henry, Subdirector Ejecutivo del Área Penal, Cibernética, de Respuesta y de Servicios del FBI aseguró en un vídeoblog que, en cualquier momento del día, alrededor de setecientos cincuenta mil hombres buscaban sexo *on line* con menores de edad.[732]

Así, por ejemplo, en 2013, la organización internacional *pro* derechos de la infancia Terre des Hommes creó un *chatbot* con apariencia hiperrealista de una niña filipina de diez años, llamada Sweetie, con la finalidad de detectar posibles pedófilos *on line*. Y es que tal herramienta (hoy en día mejorada con la versión Sweetie 3.0) cuenta con la capacidad de involucrarse y seguir las conversaciones sexualmente sugerentes que los hombres tienen con ella, registrando y almacenando toda la información al respecto, que se emplea para detectar,

[732] FBI – Federal Bureau of Investigation, 2011.

rastrear e identificar a los posibles pedófilos y, posteriormente, proceder de forma fundada a su detención y ulterior condena.[733]

No obstante, en nuestro Derecho Penal tal herramienta de IA tiene un encaje y una utilidad limitados.

Así, tal y como afirman José R. y Alejandra Vargas, que han analizado este asunto en profundidad, "*actualmente el ordenamiento jurídico español no cuenta con los específicos instrumentos legales que darían plena y pacífica cobertura jurídico-procesal a la puesta en práctica de un proyecto como Sweetie 2.0*."[734] No obstante, "*podría buscársele un encaje legal mediante dos vías que se complementarían: a través de alguna de las medidas de intervención de las comunicaciones (la más adecuada sería la prevista en el art. 588 quáter a LECrim) y, al mismo tiempo, por medio de la figura del agente encubierto virtual, online o informático del art. 282 bis 6 y 7 LECrim, aunque con la particularidad de que, en realidad, la actuación del agente de la autoridad vendría sustituida por un avatar dotado de inteligencia artificial*."[735]

En cualquier caso, tal y como señalan los mencionados autores, la utilización indiscriminada por parte de la policía española de herramientas tales como Sweetie podría implicar dos riesgos: por un lado, la vulneración del principio de especialidad ("*en tanto que se plantean serias dudas sobre si, al tratarse la supuesta víctima de un simple avatar, se cumpliría la exigencia de un delito concreto que justifique la injerencia*"[736]); y, por otro lado, la vulneración de la prohibición de realizar investigaciones prospectivas o *phishing expeditions*, "*en tanto que la utilización indiscriminada de avatares en el ciberespacio no obedecería a una sospecha razonable, objetiva y concreta*". Así, "*solo sería admisible la utilización de medidas de investigación como las propuestas en el proyecto Sweetie 2.0 cuando la finalidad fuera constatar los extremos de una concreta actividad delictiva de la que, previamente, se haya tenido noticia o sospecha razonable*."

No obstante, surge la pregunta: ¿la mera interacción de un sospechoso con una herramienta virtual como Sweetie podría considerarse una noticia o sospecha razonable para justificar, por

[733] Véase más Terre des hommes, s.f..
[734] Agustina & Vargas, 2019, pág. 617.
[735] *Idem*. Pág. 9.
[736] *Idem*.

ejemplo, el inicio de una investigación penal y, en su caso, adoptar medidas restrictivas de derechos (tales como, por ejemplo, una entrada y registo, un volcado de la información del ordenador, etc)? En mi opinión, la respuesta, en principio, es negativa.

Y es que, tal y como afirman los mencionados autores, "*a la luz de la legislación y jurisprudencia españolas, difícilmente puede sostenerse algún tipo de responsabilidad penal a partir de la sola interacción del sospechoso investigado con un avatar controlado, en última instancia, por la policía.*"[737] Y ello con base, principalmente, en dos motivos: por un lado, el hecho de que el sospechoso no tenga relación con un menor de edad real, sino con un mero *chatbot*, supone la concurrencia de un delito absolutamente imposible por inexistencia de objeto; y, por otro lado, el hecho de que el sospechoso actúe como consecuencia "*de la incitación directa o indirecta de un avatar debería entenderse, en principio, como un delito provocado, por lo que debería considerarse penalmente irrelevante, procesalmente inexistente y, por todo ello, impune*".[738]

No obstante, cierto es que por ejemplo, en aquellos casos en que la policía tuviera sospechas de que un determinado individuo formara parte de una red de explotación infantil, o de que estuviera manteniendo conversaciones con diversos menores con contenido sexual inapropiado, el uso de Sweetie para comprobar su comportamiento y su *modus operandi* podría resultar no solo legítimo sino también útil, siendo que actuaría como un agente encubierto que podría llegar incluso a extraer información del caso concreto investigado.

b) Otras herramientas

b.1) *Herramientas para detectar denuncias falsas*

b.1.1) *Concepto*

En el ámbito de la investigación criminal, también, y en concreto en España, especialmente relevante es hacer referencia a la herramienta VeriPol, pionera a nivel mundial, creada por la Policía Nacional, que con una precisión de más del 90% (frente al 75% de los agentes

[737] Agustina & Vargas, 2019, pág. 619.
[738] *Idem.*

expertos, según el Ministerio del Interior[739]) determina si una denuncia por robo con violencia e intimidación es real o es falsa.

Tal aplicación informática, que emplea técnicas de PLN, fue ideada por el inspector Miguel Camacho (matemático, además de miembro del Cuerpo Nacional de Policía), que tal y como manifestó a RTVE: "*Veíamos que había unas características comunes en ciertas denuncias, unas que podían asignarse a las que podían ser falsas y otras a las que eran verdaderas*"[740], por lo que diseñó un algoritmo capaz de detectar tales variables y calcular las probabilidades de que los denunciantes estuvieran haciendo manifestaciones falsas.

No obstante, no es VERIPOL quien toma las decisiones, sino que sirve como mera alerta y apoyo para los agentes de policía, que son quienes, en última instancia, deciden iniciar o no una investigación por presunto delito de denuncia falsa.

b.1.2) *Posibles utilidades en la instrucción de las causas*

Este tipo de herramientas, desde luego, por un lado, permite un ahorro de recursos policiales humanos, habida cuenta de que lleva a cabo un primer filtro, muy riguroso, de aquellas denuncias que pueden resultar falsas, por lo que facilita el trabajo a los agentes, lo que se traduce en una mayor agilidad en la detección de tales casos. Por otro lado, asimismo, esta clase de herramientas permitiría aportar un indicio más a las causas de instrucción seguidas por denunca falsa, si bien en cualquier caso para ostentar valor probatorio debería ser sometido a contradicción en el acto del plenario.

Así, imaginemos que llega al Juzgado de Instrucción un atestado en que se ponen de manifiesto unos hechos que podrían ser constitutivos de delito de denuncia falsa. Pensemos, asimismo, que los únicos indicios existentes son la imposibilidad de los agentes de corroborar la versión de los hechos dada por el denunciante y la sospecha de estos de que, por la forma de explicar los hechos, aquél está mintiendo. Ello, no obstante, muy posiblemente, en caso de que la persona investigada negara los hechos, podría acabar con un sobreseimiento provisional de la causa por falta de indicios racionales de criminalidad. Sin embargo, en caso de que se

[739] Ministerio del Interior, 2018.
[740] Véase Kolotushkina, 2018.

aportara el resultado arrojado por el sistema de IA de detección de denuncias falsas, siempre que pudiera ser sometido a contradicción en el acto del plenario, los indicios podrían quedar reforzados y el caso podría prosperar judicialmente.

b.2) ***Herramientas para detectar y, en su caso, moderar contenido online***

b.2.1) *Concepto*

Las técnicas de PLN, sin duda, pueden resultar muy útiles para el análisis de contenido *on line*, pudiendo detectar patrones semánticos inusuales que podrían ser indicativos de la comisión de ilícitos penales.

Y es que el PLN puede ofrecer un gran apoyo a los agentes de la autoridad, especialmente a la hora de analizar vastos contenidos en Internet lo que, de hacerse por un humano, devendría una tarea ardua, compleja e inabarcable (por ejemplo, para buscar interacciones entre miembros de una organización terrorista, para hallar posibles conversaciones dirigidas a la comisión de delitos de trata de seres humanos, para detectar mensajes que podrían ser constitutivos de delitos de explotación infantil, etc). Y es que, los contenidos de la red, por su enorme amplitud, son imposibles de examinar de forma exhaustiva si no es con ayuda de una "máquina", habida cuenta de que los recursos humanos son evidentemente limitados.

En relación con ello, procede poner de manifiesto que, por ejemplo, los autores de los tiroteos ocurridos en la mezquita en Christchurch (Nueva Zelanda) el 15 de marzo de 2019), y de la sinagoga de Poway (California, EEUU) el 27 de abril de 2019 realizaron publicaciones en la página web "8chan" antes de cometer sus ataques terroristas.

Y es que se entiende que con un sistema de PLN potente se podría, sin duda, identificar el incremento de los niveles de amenaza, debiendo adaptarse los indicadores a las características específicas de la plataforma y de la legislación nacional en que este operara, siempre con respeto a los derechos fundamentales de los ciudadanos.

Las voces críticas, no obstante, apoyan lo que podría considerarse una interpretación extrema del valor de la libertad de expresión y, por lo tanto, entienden que el uso de este tipo de herramientas podría limitar los derechos fundamentales de los usuarios, por lo que deberían

quedar prohibidas. En mi opinión, ello es cierto, sin embargo, con una buena regulación que fijara los límites, tales medidas podrían respetar las normas de privacidad y los derechos humanos en todo momento y ser empleadas con fines legítimos, de forma muy útil.

b.2.2) *Posibles utilidades en la instrucción de las causas*

Ante la amenaza yihadista que acecha a Europa, cuerpos policiales de los distintos países destinan un gran número de agentes para realizar tareas de rastreo de la red con el fin de detectar posibles casos de captación de adeptos y radicalización, publicación de contenido terrorista o preparación de ataques, entre otros. Como consecuencia de ello, un ingente número de personas se dedican en cuerpo y alma, durante largas jornadas de trabajo, a analizar millones de páginas web y a tratar de establecer conexiones y detectar patrones irregulares, lo cual no solo supone un coste extraordinario en recursos materiales y humanos, sino que también implica la existencia de limitaciones en la investigación, habida cuenta de que dichos recursos no son infinitos.

No obstante, mediante la utilización de técnicas automáticas de PLN, el trabajo que miles de policías realizan a diario podría ser, en parte, sustituido y, desde luego, mejorado. Y es que tales sistemas permitirían que los agentes se dedicaran simplemente a verificar, una vez el sistema de IA hubiera hecho saltar la alarma, qué es lo que ocurre realmente y si, en su caso, tiene relevancia a efectos de investigación penal.

Además, hay que tener en cuenta que en los cuerpos policiales se cuenta con agentes que tienen conocimientos de distintos idiomas, pero obviamente tales conocimientos suelen ser de lenguas mayoritarias y, solo en casos muy concretos, se cuenta con especialistas en idiomas minoritarios. No obstante, mediante el uso de sistemas de PLN, el abanico de posibilidades se amplía de forma significativa, habida cuenta de que estos pueden ser empleados para analizar contenidos realizados en lenguas residuales que, de otra forma, pasarían desapercibidos, lo que podría ser empleado por los presuntos delincuentes para llevar a cabo de forma impune su actividad delictiva.

b.3) ***Herramientas para analizar documentos***

b.3.1) *Concepto*

A modo de concepto, los sistemas de IA de análisis de documentos son aquellas herramientas que, por lo general, a partir de técnicas de PLN (combinadas con técnicas de aprendizaje automático o *Machine Learning*, entre otras), tienen capacidad para examinar ingentes cantidades de datos e interpretarlos. Así, dichos sistemas pueden detectar patrones, filtrar y clasificar conceptos clave, categorizar elementos, relacionar informaciones, etc.

Ello, sin duda, puede resultar de gran ayuda para llevar a cabo el examen de ingentes cantidades de documentos con contenido vasto y diverso, lo que desde luego supone un auténtico reto (muchas veces prácticamente inabarcable) para los humanos, que vemos cómo nuestro tiempo se consume al hacer profundas inmersiones documentales que no siempre acaban de forma satisfactoria, dadas las limitaciones intelectuales, materiales y temporales de que disponemos.

No obstante, la buena noticia es que este tipo de sistemas ya ha llegado a las Administraciones de Justicia de algunos países del mundo, lo cual resulta ciertamente alentador, habida cuenta de la enorme cantidad de documentación que se presenta en los juzgados para su posterior análisis.

En relación con ello, y como paradigma del uso y la utilidad de dicha tecnología, procede hacer referencia al denominado "Caso Rolls-Royce"[741], resuelto ante la justicia británica después de que dicha compañía reportara a la oficina antrifraude inglesa ("*Serious Fraud Office*" –SFO-) que algunos de sus intermediarios habían realizado sobornos en países como Indonesia, China y Brasil, entre otros.

Y es que ya en 2012 las autoridades británicas iniciaron una investigación sobre las presuntas malas prácticas de la mencionada compañía y, posteriormente, en 2016 diversos medios de comunicación británicos publicaron testimonios y documentos que insinuaban que Rolls-Royce había empleado una red internacional de contactos para conseguir suculentos contratos a través de sobornos. Tras ello, la propia firma admitió los hechos y su entonces Consejero

[741] Véase Courts and Tribunals Judiciary, 2017.

Delegado, John Rishton, aseguró que su voluntad era la de eliminar cualquier negocio indebido y colaborar con las autoridades.

Como consecuencia de ello, el 17 de enero de 2017 Rolls-Royce anunció que pagaría seiscientos setenta y un millones de libras (unos setecientos sesenta millones de euros) a las autoridades del Reino Unido, Brasil y Estados Unidos para resolver el caso.[742]

No obstante, en el marco de tal investigación se produjo uno de los mayores hitos tecnológicos de los últimos años en el ámbito de la Administración: el del uso de sistemas de IA de análisis automático de documentos, lo que permitió a la SFO investigar de forma más rápida, con menores costes y una menor tasa de error que en el caso de haber empleado letrados humanos.

Y es que tal y como se pone de manifiesto desde la propia SFO, en tal caso se empleó un sistema de IA piloto capaz de procesar más de medio millón de documentos al día (de un total de treinta millones de documentos) a velocidades dos mil veces más rápidas que un abogado humano.

Tras el éxito de tal innovación, la mencionada oficina antifraude anunció la incorporación de "*Open Text Axcelerate*", un sistema de IA de revisión de documentos flexible y potente construido a través de a la automatización, el aprendizaje automático y la analítica avanzada patentada y que, según se anuncia en su página web "*ofrece las mejores capacidades de investigación de su clase en una interfaz de revisión intuitiva y totalmente integrada que ayuda a los equipos legales a llegar antes a los hechos importantes y decidir la estrategia del caso*."[743] Y es que tal sistema, tal y como se anunció por la propia SFO, no solo puede reconocer patrones, agrupar información por temas, crear líneas temporales y eliminar duplicados, sino que además puede analizar la relevancia y llegar a eliminar aquellos documentos que no estén relacionados con una investigación.

En relación con ello, el Director de tecnología de la SFO, Ben Denison, aseguró: "*La tecnología de IA nos ayudará a trabajar de manera más inteligente, rápida y eficaz para investigar y enjuiciar los delitos económicos. El uso de tecnología innovadora como esta ya*

[742] Véase ABC, 2017.
[743] Opentext, s.f..

no es opcional, es esencial dado el volumen de material con el que estamos tratando y ayudará a garantizar que podamos seguir cumpliendo con nuestras obligaciones de divulgación y hacer justicia de forma más rápida y a un coste significativamente menor."[744]

Así, el primer caso que la SFO analizó con dicho novedoso sistema de IA superaba ya al caso Rolls-Royce en relación con el número de documentos a revisar, con más de cincuenta millones. Y es que tal herramienta, sin duda, puede servir de enorme ayuda a los investigadores para analizar y procesar cantidades masivas de información con el fin de detectar la verdad que subyace entre millones de datos. En concreto, "*Open Text Axcelerate*" permite a día de hoy la revisión de más de cien mil documentos al día, lo cual es absolutamente abrumador en comparación con la capacidad de un ser humano de revisar tan solo unos pocos de cientos de documentos en el mismo lapso de tiempo.

Ello, tal y como apunta Richard Day, Jefe de eDiscovery, de la SFO, permite "*a los investigadores usar más su inteligencia, hacer el trabajo más duro, el trabajo más complejo, el trabajo que disfrutan.*".[745]

b.3.2) *Posibles utilidades en la instrucción de las causas*

En virtud de lo expuesto, sin duda alguna entiendo que este tipo de sistemas puede resultar de enorme utilidad para los investigadores en el seno de los procesos penales (tanto en sede policial como judicial). Así, ello podría otorgar un gran apoyo a los investigadores especialmente en aquellos casos en que deben ser examinadas ingentes cantidades de documentos, lo cual supone un enorme consumo de recursos personales, genera un gran coste de oportunidad (habida cuenta de que, mientras estos están examinando documentos, están dejando de destinar su tiempo y sus capacidades a otros asuntos también relevantes) y puede hacer tambalear la calidad de las investigaciones, siendo que muchas veces es materialmente imposible llevar a cabo un análisis exhaustivo de los cientos de documentos que se aportan como prueba.

Y es que, imaginemos que llega a oídos de la policía la existencia de una organización criminal que está llevando a cabo una trama de delitos de estafa y blanqueo de capitales y, como

[744] Serious Fraud Office, 2018.
[745] Linscott, 2020.

consecuencia de ello, se decide iniciar una investigación. Pensemos, asimismo, que tras recopilar una serie de indicios, la policía solicita al juez de guardia diversas entradas y registros simultáneas en varios domicilios y, una vez allí, los agentes incautan decenas de ordenadores cuyo contenido deberá luego ser analizado (tras la realización del correspondiente volcado). Supongamos, asimismo, que dentro de los ordenadores se hallan cientos de miles de correos electrónicos cruzados entre los distintos miembros de la trama que, para el buen fin de la investigación, deben ser exhaustivamente analizados.

Ante tal circunstancia, en caso de no contar con un potente sistema de IA que permitiera examinar toda la información incautada, filtrarla y ordenarla para facilitar el trabajo a los agentes, estos deberían pasar semanas, incluso meses, dedicando el 100% de su jornada laboral a analizar el contenido de los correos electrónicos, lo cual es altamente ineficiente en todos los sentidos. No obstante, mediante el uso de un buen sistema de IA de análisis documental, dicho trabajo de los agentes podría resultar altamente simplificado, ya que este les proporcionaría el contenido filtrado, interpretado y sistematizado, lo que les serviría de guía para proceder a extraer y analizar únicamente aquella información que les resultara útil y necesaria, ahorrándose un trabajo previo altamente arduo y demandante.

Y es que el sistema de IA de análisis de documentos podría, sin duda, aportar en cuestión de horas (o días, según el volumen de la información a examinar) análisis e interpretaciones de la información llevados a cabo desde distintos puntos de vista, a saber, por ejemplo: agrupación de todos los correos electrónicos enviados y recibidos por una/s concreta/s persona/s; clasificación de los correos electrónicos por meses o por años; relación y agrupación de aquellos correos electrónicos que trataran sobre los mismos temas; agrupación de aquellos correos electrónicos que llevaran documentos adjuntos; relación de aquellos correos electrónicos que contuvieran ciertas palabras clave; eliminación de aquellos correos electrónicos que no tuvieran relación alguna con los hechos investigados, reduciendo así significativamente la cantidad de información a tener en cuenta, etc.

Y ello, desde luego, permitiría a los agentes contar de forma rápida y efectiva con una fotografía mental del contenido de los cientos de miles de correos electrónicos que, de otra forma, hubieran tardado meses en analizar, lo que hubiera ido, sin duda, en detrimento de la

investigación, que especialmente después de una diligencia de entrada y registro requiere de cierta celeridad para garantizar su éxito.

Y lo expuesto, desde luego, aplica igualmente en el caso de que una investigación penal se halle completamente judicializada, por ejemplo, en virtud de la admisión a trámite de una querella que contiene, como prueba documental, miles de documentos que deben ser analizados por el Ministerio Fiscal, la defensa, el juez de instrucción y, posteriormente, los magistrados del plenario.

Así, imaginemos que se interpone ante un Juzgado una querella por corrupción política que va acompañada de miles de documentos aportados en concepto de prueba. Recordemos, asimismo, que ya de entrada el juez de instrucción para decidir sobre la admisión a trámite o no de tal querella debe hacer un análisis previo del relato de hechos y la documentación aportada y, posteriormente, en caso de que la querella prospere, deberá examinar tal información con mucha más profundidad, lo cual, desde luego, es absolutamente inabarcable, máxime siendo que los juzgados están extremadamente colapsados y el número de casos a los que se enfrenta cada juez suele ser superior al que cualquier mente humana puede asumir en el corto o medio plazo.

En virtud de lo expuesto, si solo se cuenta con el elemento humano para llevar a cabo el análisis de toda la documentación aportada, lo más probable es que la causa se quede "a la cola" de los quehaceres del Ministerio Fiscal y del juez de instrucción (que no pueden permitirse parar su actividad durante días o incluso semanas o meses para analizar toda la información aportada en una sola causa), con el correspondiente retraso en la investigación y las fatales consecuencias materiales y procesales que ello puede acarrear (dilaciones indebidas, sustracción a la justicia de personas clave en la investigación, dificultades de los testigos para recordar los hechos con precisión, destrucción de pruebas, etc). Y ello es una realidad, y a pesar de que a ciertos juzgados que cuentan con alguna de las denominadas "macrocausas" se les dota de un juez de refuerzo con dedicación exclusiva a esta (o sustitución del juez titular en su trabajo ordinario a fin de que este pueda destinar el 100% de su tiempo a la mencionada investigación), lo cierto es que, por un lado, ello no siempre ocurre y, por otro lado, incluso cuando ello ocurre, las ineficiencias son patentes y la calidad de las investigaciones no siempre resulta la más deseable, puesto que una sola mente humana (sin

formación especializada, además) no es capaz de analizar, interpretar y sistematizar tanta información. Y tales ineficiencias se dan incluso en el caso de que el juez instructor se apoye en profesionales especializados (generalmente agentes de policía que cuentan con conocimientos específicos) para llevar a cabo las tareas de examen de los documentos, habida cuenta de que, por un lado, los tiempos que se manejan siempre van a ser excesivos y, por otro lado, al final el que tiene que realizar un análisis exhaustivo de la documentación obrante en autos y comprender su contenido es Su Señoría.

No obstante, si se contara con programas de IA que en este caso llevaran a cabo un filtrado previo de la información y proporcionaran al Ministerio Fiscal, la defensa y el juez una filtración previa de esta, en los términos anteriormente expuestos, las dificultades antedichas tenderían a reducirse en gran medida.

b.4) ***Herramientas de traducción simultána***

b.4.1) *Concepto*

En numerosas ocasiones las diligencias de guardia se retrasan porque el intérprete no llega o bien porque hasta el último momento no se tiene constancia de que un testigo no habla correctamente español. Y lo mismo ocurre con las diligencias de investigación ordinarias. Además, el coste del servicio de interpretación es elevado para la Administración y, no solo eso, si no que a veces ciertos intérpretes, en vez de traducir de forma simultánea, optan por resumir lo que dice la persona extranjera después de que esta hable durante un rato, lo cual puede dar lugar a la subjetividad y puede resultar peligroso, ya que lo justo es que se traduzca palabra por palabra lo que el sujeto manifiesta ante el juez de instrucción, aunque ello lleve más tiempo.

Hoy en día, no obstante, ya existen sistemas (por ejemplo, Google Translate) que, mediante el uso de la IA, son capaces de traducir de forma simultánea y automática lo que una persona dice en numerosos idiomas, lo cual es, sin duda, un gran avance. Y es que si bien estos sistemas en un principio tenían fallos patentes, hoy en día cada vez son mejores (aunque les queda camino por recorrer para alcanzar la perfección) y más fieles a la realidad.

Así, en mi opinión, a medida que esta tecnología vaya avanzando (bajo mi punto de vista, los sistemas actuales todavía no son lo suficientemente precisos como para ser empleados en una materia tan sensible como es la investigación penal), debería ir implementándose en los Juzgados, habida cuenta de que incrementaría la eficiencia, puesto que la tarea de traducir las declaraciones devendría más rápida, más objetiva y más barata.

b.4.2) *Posibles utilidades en la instrucción de las causas*

En virtud de lo expuesto, la utilidad de los sistemas de traducción e interpretación mediante IA podría ser enorme.

Y es que imaginemos que un sábado debe declarar ante el Juzgado de Guardia una detenida que habla urdú y otra que habla suajili y tras llamar en repetidas ocasiones al servicio de guardia de intérpretes, desde este no contestan o informan de que se van a poner en contacto con dos intérpretes de dichos idiomas que, no obstante, permanecen durante un buen rato ilocalizables. Y ello, aunque parezca poco probable, por desgracia es muy frecuente en el día a día de los juzgados. En tal caso, sin duda, la práctica de la diligencia de guardia se vería retrasada hasta que los mencionados intérpretes pudieran ser localizados, aceptaran el servicio y acudieran a la sede judicial (aunque cierto es que actualmente están empezando a proliferar las interpretaciones telefónicas). No obstante, en caso de contar con un sistema de IA de traducción e interpretación simultánea, tales tiempos se verían acortados y la eficiencia sin duda resultaría incrementada.

b.5) ***Herramientas de transcripción automática***

b.5.1) *Concepto*

Hace unos años, lo más frecuente era que las declaraciones efectuadas en sede judicial fueran transcritas a mano por el/la Letrado/a de la Administración de Justicia o los funcionarios de Justicia, que debían pasar largas horas en Sala recogiendo de forma minuciosa todo lo que se decía ante el juez. Ello, además de agotador para el que realizaba la tarea, suponía un retraso considerable en la práctica de las diligencias de investigación, habida cuenta de que las

declaraciones se alargaban considerablemente, puesto que había que dar tiempo a que fueran recogidas a mano (o a máquina) por el personal del Juzgado.

No obstante, hoy en día, bajo mi punto de vista, en virtud de lo dispuesto en los artículos 453 y 454 de la LOPJ (supletoria de la LECrim), en relación con lo establecido en los artículos 397, 402, 437, 443, 444 y 450 de la LECrim, el/la Letrado/a de la Administracion de Justicia tiene la facultad de decidir qué medio se emplea para dejar constancia de las declaraciones, a saber: el de escribir a mano o a máquina o el de utilizar los medios técnicos, audiovisuales e informáticos de documentación con que cuente la unidad donde presta sus servicios, lo cual está obligado/a a promover. No obstante, ello no está tan claro a la luz de lo dispuesto en el artículo 230 LOPJ, que dispone que: "*Las actuaciones orales y vistas grabadas y documentadas en soporte digital no podrán transcribirse, salvo los casos expresamente previstos en la ley*", lo cual puede da lugar a diversas interpretaciones.

En virtud de ello, no obstante, surgen en ocasiones controversias entre Letrados de la Administración de Justicia, jueces, fiscales y partes personadas, habida cuenta de que algunos reclaman la transcripción de las grabaciones, otros consideran que estas no son necesarias, etc y, de hecho, no han sido pocos los tribunales que se han tenido que pronunciar al respecto[746], en ocasiones en sentido contradictorio, lo cual no hace más que dar sentido a la implementación de un sistema de IA capaz de transcribir las declaraciones judiciales.

Y es que lo que esta claro es que lo ideal es disponer de las declaraciones tanto grabadas como transcritas, puesto que ello, por un lado, permite revisar la realidad de la diligencia tal cual ocurrió (lo cual es muy útil no solo para el propio instructor, fiscal o Letrado que las practicó sino también para el instructor, fiscal o Letrado que se hace cargo del caso con posterioridad); y, por otro lado, permite realizar una revisión más rápida y encontrar la información de modo más eficaz.

Y ello, con el uso de sistemas de IA, es posible. Así, podría resultar preceptiva la grabación de las declaraciones, por un lado, y, por otro, la transcripción de las mismas a través de tal tipo de herramientas, lo cual proporcionaría a los fiscales, jueces y demás operadores jurídicos

[746] Entre otros, véanse Auto de la Audiencia Provincial de Sevilla de 1 de febrero de 2019 y Auto de la Audiencia Provincial de Barcelona de 8 de julio, de 2019.

la máxima información posible y garantizaría más si cabe los derechos de las partes implicadas. Y es que no son pocas las ocasiones en las que se ha perdido una declaración que solo constaba escrita en papel o se ha dañado una declaración que solo constaba grabada, lo cual genera inconvenientes graves (y, en ocasiones, puede incluso hacer fracasar la instrucción).

En España, el Ministerio de Justicia ya está llevando a cabo un proyecto de transcripción automática con IA en el que ya se han transcrito más de veintitrés mil grabaciones de vistas judiciales. El sistema, comercializado por la compañía española Shooowit Stream, S.L. no solo permite transcribir las declaraciones de las partes al momento sino también guardarlas y descargarlas en formato "Word" o "pdf". Y es que, con el fin de facilitar el trabajo a todos los operadores jurídicos, el objetivo es que con el mencionado sistema cada grabación, además, contenga la transcripción en subtítulos sobre la pantalla, para verla a tiempo real, y en una columna lateral. Además, dicho sistema ofrece la posibilidad de buscar por palabras y por frases; incluir anotaciones; segmentarlo en secciones; etiquetarlo y marcar los apartados más relevantes.[747]

Bajo mi punto de vista, si bien ello es una buena noticia, puesto que significa un muy esperado avance en la buena dirección, todavía queda mucho camino por recorrer. Y es que habrá que ver el nivel de precisión con que cuenta dicho sistema, que además por el momento únicamente está disponible para el idioma español, los problemas que luego pueda acarrear en su aplicación real, etc.

b.5.2) *Posibles utilidades en la instrucción de las causas*

En virtud de lo expuesto, tales sistemas podrían aportar una gran utilidad en la instrucción de las causas (y también en fase de plenario).

Y es que, por un lado, en el ámbito policial (con incidencia en la instrucción), imaginemos una causa judicializada en la que se investiga a una organización criminal por la presunta comisión de un delito contra la salud pública y el juez de instrucción ha acordado la

[747] Véase Berbell, 2020.

intervención de más de una veintena de líneas telefónicas y la instalación de dispositivos de escucha en una decena de vehículos. Pensemos, asimismo, que las tareas de la mencionada organización van muy rápidas y que prácticamente a diario se requiere de una actualización al juez instructor para que acuerde nuevas diligencias de investigación con el fin de detectar y, en su caso, frustrar, la comisión de nuevos hechos.

En tal supuesto, si la única forma de que el contenido de las conversaciones captadas resultara trasladado al juez fuera la transcripción a mano efectuada por los agentes de policía encargados del caso, el despliegue de recursos sería absolutamente desproporcionado. Sin embargo, en caso de poder emplear un sistema de IA de transcripción automática (que, incluso, combinado con un sistema de reconocimiento de voz permitiría identificar a quién corresponde la voz de cada hablante en cada momento), el uso de recursos resultaría muchísimo más eficiente y sostenible (y, de hecho, con toda seguridad la tarea se efectuaría en un lapso de tiempo récord), lo que permitiría a los agentes ocuparse de aspectos que solo los humanos, por el momento, pueden gestionar (establecer relaciones y conexiones, interpretar, sacar conclusiones más profundas, etc), y al juez de instrucción acelerar la práctica de sus diligencias.

Por otro lado, en el ámbito judicial propiamente dicho, imaginemos una causa de estafa piramidal en la que hay diversos perjudicados, testigos e investigados, cuyas declaraciones duran una media de media hora cada una, habida cuenta de la complejidad del asunto. Pensemos que la instrucción de tal caso dura más de un año, periodo en el que el juez de instrucción se cambia de juzgado y es sustituido por uno nuevo, que deberá ponerse al día de todo para continuar con la dirección de la investigación.

En tal supuesto, si las declaraciones únicamente constaran en un soporte apto para la grabación y la reproducción del sonido y de la imagen, la tarea para el juez de instrucción entrante a la hora de dictar Auto de continuación de Procedimiento Abreviado o de sobreseimiento, y para los Letrados y el Ministerio Fiscal, en su caso, para preparar sus recursos o escritos de calificación y defensa, sería ciertamente ardua y consumiría una gran cantidad de su tiempo, habida cuenta de que deberían visualizar una a una las mencionadas declaraciones. No obstante, en caso de que estas solamente constaran por escrito, en papel o en formato digital, sin duda se perdería una gran parte del contenido de las mismas, habida cuenta de que también

es importante analizar el lenguaje no corporal (y es que, bajo mi punto de vista, la única forma de que el juez salvaguarde el principio de inmediación es que presencie o vea una declaración grabada). Y, finalmente, en caso de que las declaraciones constaran tanto grabadas como escritas, es evidente que la cantidad de tiempo que su práctica habría llevado hubiera sido excesiva.

Así, como ya se ha expuesto, lo más idóneo es contar con las grabaciones tanto grabadas como transcritas y, ello, con un sistema de IA que permitiera grabar y transcribir de forma simultánea, resultaría perfectamente viable.

B.3. *Regulación española y europea*

Respecto de ello, y en aras de evitar duplicidades, siendo que en la mayoría de los sistemas analizados se emplean datos de carácter personal, procede remitirse a lo dispuesto, con carácter general, en relación a las herramientas de IA que emplean datos biométricos (habida cuenta de que estos tienen la consideración de datos de carácter personal) y, por ende, todo lo que les resulta aplicable es extensible a las herramientas que ahora nos ocupan, con la salvedad de lo dispuesto de forma específica para tal clase de sistemas.[748]

Así, procede hacer mención, de forma principal, a lo dispuesto en la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales y a lo previsto en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión publicada por la Comisión Europea el 21 de abril de 2021, si bien tal y como ya se ha venido poniendo de manifiesto ello no es más que una mera propuesta, como su propio nombre indica, y por ende, debe tomarse con las oportunas cautelas.

B.4. *Riesgos jurídicos generales*

a) Posible falta de precisión y potencial discriminatorio

[748] Véanse págs. 416-440.

Por un lado, si bien algunos de los sistemas que emplean técnicas de PNL cuentan con elevados grados de precisión, como es el caso de VeriPol, que según fuentes del Ministerio del Interior dispone de una tasa de acierto de más del 90% a la hora de estimar la probabilidad de que una denuncia por robo con violencia e intimidación o tirón sea falsa[749], lo cierto es que ello no ocurre en con la totalidad de este tipo de sistemas. Y no solo eso, sino que en la mayoría de ocasiones, incluso en aquellos casos en que se asegura que las tasas de éxito son elevadas, no se explica realmente cómo se ha llegado a tales conclusiones, lo cual, cuanto menos, resulta inquietante.

Y la precisión en este tipo de sistemas (como en todos los que emplean IA, en especial en el ámbito de la investigación penal) resulta absolutamente fundamental, habida cuenta de que un solo fallo puede conllevar nefastas consecuencias.

Así, imaginemos que una herramienta de análisis documental comete un error al eliminar un documento por entender que no tiene relevancia para la causa y, sin embargo, sí la tiene. O imaginemos que un *chatbot* no detecta que una víctima de violencia de género le está dando la orden de llamar a la policía o le genera de forma automática respuestas erróneas a sus preguntas sobre los pasos a seguir. O que una herramienta de moderación de contenido *on line* no capta una amenaza terrorista. O, que una traducción de escasa calidad conlleve consecuencias absolutamente nefastas para la persona que habla el idioma extranjero, habida cuenta de que en el ámbito de la investigación criminal "el diablo está en los detalles" y todo cuenta. Y es que en tal caso de no resultar bien traducida una declaración, tanto el sujeto que la presta como aquellos que deben soportar las consecuencias de la misma pueden ser puestos en una posición de indefensión total y absoluta, ya que salvo que el Letrado, el Ministerio Fiscal o el juez perciban algo extraño y decidan consultar a un intérprete humano, lo normal es que den por buena la declaración, lo cual podría llevar a tomar decisiones judiciales erróneas. Y es que ¿cuál es la forma de comprobar que lo traducido por el sistema es exactamente lo declarado por un sujeto de habla extranjera? El propio declarante resta en posición de impotencia total al no entender el español y no poder verificar si lo que se refleja

[749] Véase Ministerio del Interior, 2018.

como declarado por él es así exactamente; el Letrado se halla en la misma situación; y desde luego las autoridades policiales, fiscales y judiciales también.

Una posible solución a este último caso, no obstante, siempre podría ser que en fase de segunda instancia los Letrados y el Ministerio Fiscal pudieran solicitar la revisión humana de la traducción efectuada por una máquina, si bien entiendo que en cualquier caso habría que dar un motivo claro, excepcional y justificado para que ello pudiera ser admitido por el órgano judicial que debería examinarse caso por caso con el fin de no colapsar la Administración de Justicia.

En virtud de lo expuesto, pues, habría que asegurarse de que este tipo de sistemas no contiene prácticamente fallos y es preciso casi al 100%.

De acuerdo con ello, los fabricantes y comercializadores de este tipo de herramientas deberían asegurar al máximo, de forma transparente, cuál es su verdadero grado de precisión y cuál ha sido el proceso que les ha llevado a sacar conclusiones al respecto, principalmente para que los usuarios puedan contar con toda la información y decidir así sobre su posible uso. Y es que justamente la Administración debe tener especial cuidado a la hora de autorizar la utilización de herramientas tecnológicas, por lo que los estándares requeridos a las compañías privadas que los diseñan y los distribuyen son mayores.

Por otro lado, procede poner de manifiesto que las herramientas que emplean técnicas de PLN no son igual de precisas en todos los idiomas. Así, siendo que los algoritmos suelen ser entrenados para comprender y analizar las lenguas mayoritarias, cuentan con mayores tasas de éxito en aquellos idiomas más generalizados, y con grados de precisión menores en aquellos cuyo uso es más limitado.

Y es que a las grandes empresas privadas que suelen destinar ingentes cantidades de dinero en analizar y filtrar sus contenidos (tipo Facebook, Instagram, etc) no suele salirles lo suficientemente rentable entrenar a los algoritmos para comprender lenguas habladas por minorías, por lo que estas caen en el olvido, y ello, sin duda, puede generar grandes diferencias que pueden conllevar resultados discriminatorios.

Así, una víctima de violencia de género que hable un idioma minoritario no tendrá la misma oportunidad de usar un *chatbot* para conectar con la policía que otra que hable un idioma mayoritario; y un ciudadano que presente documentación escrita en un idioma minoritario deberá esperar a que su caso sea resuelto en veinte veces más tiempo que uno que presente documentación escrita en uno mayoritario.

Y no solo eso, sino que las diferencias existentes en cuanto a las distintas lenguas empleadas pueden también incidir en el grado de precisión de los sistemas, puesto que, desde luego, una herramienta que emplee técnicas de PLN captará más rápidamente una amenaza terrorista *on line* lanzada en inglés que otra lanzada en un dialecto del urdú.[750]

Estos sistemas, pues, deberían ser introducidos de la forma más uniforme y equitativa posible, con el fin de intentar ofrecer la misma calidad del servicio para los idiomas mayoritarios y los minoritarios y evitar así potenciales discriminaciones. Y es que no es justo que una persona, por hablar una lengua minoritaria, no pueda tener acceso a un sistema de IA de traducción simultánea, con los beneficios que ello puede conllevar, por lo que habría que asegurar que la inversión en este tipo de sistemas se hiciera en relación a todas o prácticamente todas las lenguas del mundo (o, al menos, habría que tener localizadas a compañías locales de cada país que pudieran dar acceso inmediato a sistemas que funcionaran con lenguas minoritarias).

b) Posible vulneración del derecho a la privacidad y a la protección de datos personales

Hay que tener en cuenta que las herramientas de IA que emplean técnicas de PLN manejan muchísimos datos, gran parte de ellos personales y, por ende, merecedores de especial protección, en virtud de lo dispuesto en el RGPD y, en concreto, en aplicación de lo previsto en la LO 7/2021, de 26 de mayo, de de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales.

Y es que los *chatbots*, por ejemplo, siendo que tienen como finalidad establecer conversaciones entre un humano y la máquina y ejecutar órdenes, con toda seguridad

[750] Idioma nacional de Pakistán.

necesitarán obtener información de su interlocutor para llevar a cabo la interacción, con el objetivo de dar la mejor y más óptima respuesta a sus necesidades.

Por su parte, los sistemas de análisis documental, sin duda tienen acceso y tratan numerosas informaciones que pueden ser calificadas como datos de carácter personal, siendo que conforman justamente el contenido de todos aquellos documentos sometidos a examen.

Y lo mismo los sistemas de moderación de contenido *on line*, que al rastrear la red tratan millones de datos, muchos de ellos de carácter personal.

De acuerdo con ello, y con el fin de que puedan ser empleadas de forma legítima y exitosa, resulta fundamental que den cumplimiento a la normativa aplicable en materia de protección de datos ya analizada en secciones anteriores (y, cuando sea publicada, asimismo deberán cumplir con la normativa específica en materia de IA).

Y es que en caso contrario podrían dar lugar a pruebas útiles pero ilegales, o a mecanismos de apoyo a las autoridades investigadoras que viciarían de nulidad los procesos de instrucción, con todas las indeseables consecuencias que ello podría conllevar.

c) Posibles brechas de seguridad

Tal y como se ha expuesto, este tipo de herramientas tratan datos personales y, sobre todo, gestionan información altamente sensible, merecedora de los máximos niveles de protección y cuidado.

Así, este tipo de herramientas debe cumplir con unos estándares de seguridad muy elevados y con unos sistemas de control de que ese nivel de seguridad se mantiene en el tiempo con la mayor calidad, para lo cual debería darse cumplimiento a lo normativa en materia de seguridad vigente y, en su caso, futura, ya analizada en secciones anteriores.

Y es que imaginemos, por ejemplo, que un *hacker* puede acceder al contenido de un expediente que está siendo analizado por una herramienta de análisis documental y se halla situación de secreto de sumario. Pensemos, asimismo, que ese hacker ha sido contratado por

uno de los implicados en la causa, que ha recibido un chivatazo y ha decidido contratar a un experto informático para que ataque el sistema y borre y manipule documentación, lo que, sin duda, podría generar consecuencias nefastas para la investigación en curso.

O, asimismo, imaginemos que se lleva a cabo un ataque informático a un *chatbot* creado por la policía para comunicarse con las víctimas de violencia doméstica y este deja de funcionar en el preciso momento en que una de ellas necesita ayuda y la deja completamente desprotegida.

O incluso pensemos en un ataque a un sistema de moderación de contenido *on line* que sustrae todo aquel contenido que se marca como potencial pelirgo y/o se pretende eliminar por la agencia de policía que lo está usando, poniendo en riesgo la seguridad nacional.

d) Posible falta de transparencia

Y, tal y como se ha puesto ya de manifiesto en relación con el resto de herramientas de IA, el requisito fundamental para poder garantizar su control y su éxito, en este caso también, es el de la transparencia. Y es que el hecho de que un sistema sea transparente (y explicable) es la única forma de verificar que cumple con todos los estándares y requisitos legalmente previstos.

Así, la eventual falta de transparencia resulta un riesgo especialmente grave en relación a este tipo de herramientas, habida cuenta de que, al igual que ocurre con las que emplean datos biométricos, en muchas ocasiones sus resultados podrían ser aportados como prueba al proceso penal y, por ende, deberían poder ser objeto de contradicción por el Ministerio Fiscal y las partes personadas; y, asimismo, algunos de los mencionados sistemas podrían servir de apoyo a las autoridades policiales, fiscales y judiciales en sus labores de investigación, lo que implica que su contenido y su funcionamiento deberían ser públicos y auditables de forma sencilla y eficaz.

Respecto de ello, imaginemos que un juez dicta un auto de sobreseimiento provisional en un caso de estafa con base en el contenido de una gran cantidad de documentos examinados por un sistema de IA de análisis documental que ha decidido eliminar cientos de ellos por

considerar que no tienen relación con los hechos. Pensemos, asimismo, que la acusación particular recurre la resolución de archivo y alega que no se han tenido en cuenta documentos que eran de especial relevancia para la causa, solicitando al juez de instrucción que exponga cuál fue el motivo que le llevó a obviarlos. Figurémonos, además, que, ante tal petición, el juez de instrucción simplemente alega que delegó la tarea de filtrar la documentación aportada en un sistema de IA y que confió en su resultado, sin dar acceso a la acusación ni a su contenido ni al modo de funcionar para tomar decisiones (tales como la de eliminar documentación). Ello, sin duda, vulneraría el derecho fundamental a la tutela judicial efectiva y a la defensa de la parte que ha visto limitados sus derechos por una decisión de un sistema de IA empleado por un juez que ni siquiera explica qué hay detrás del mismo, qué criterios tiene en cuenta, qué grado de precisión maneja, etc lo cual resultaría absolutamente intolerable.

C- Herramientas de IA que emplean técnicas de Visión Artificial o *Computer Vision*

C.1. *Concepto*

Las técnicas de Visión Artificil o *Computer Vision* se caracterizan por tener la capacidad de analizar, interpretar y extraer información a partir de imágenes digitalizadas formadas por mapas de píxeles.

Así, las herramientas de IA que emplean este tipo de ténicas, que se nutren de algoritmos de *Machine Learning* y *Deep Learning*, pueden resultar de enorme utilidad para llevar a cabo análisis automáticos masivos o particulares de imágenes que constan en formato digital, lo cual sin duda puede revertir en beneficio de la investigación criminal.

C.2. *Subclases*

a) Herramientas de análisis de imágenes

a.1) *Concepto*

A modo de concepto, el reconocimiento de imágenes es aquella tecnología que, a través de la

IA (y, en concreto, de técnicas de Visión Artificial o *Computer Vision*), procesa ciertas figuras y/o símbolos y reconoce la información contenida en ellos.

Las técnicas de análisis de video e imágenes en la actualidad ya son empleadas por las autoridades policiales para obtener información sobre personas, objetos y acciones de interés para la investigación penal, si bien requieren muchos recursos personales, lo que requiere grandes inversiones en agentes especializados y con experiencia.

No obstante, es interesante poner de manifiesto que a las máquinas se les puede enseñar a analizar e interpretar imágenes de la misma forma que lo hacen nuestros cerebros y ello, combinado con el enorme potencial que estas tienen, por su acceso a ingentes cantidades de datos y por su capacidad para examinarlos a velocidades humanamente imbatibles, puede resultar muy útil para la investigación criminal.

En términos generales, el procesamiento o el análisis de imágenes, en el ámbito de la investigación penal, consiste en examinar y tratar una imagen digitalizada para extraer información de ella. Y es que los algoritmos no solo son capaces de aprender tareas complejas, sino que asimismo pueden desarrollar y determinar sus propios parámetros de análisis que van más allá de lo que los humanos pueden conseguir, siendo capaces de identificar armas y otros objetos y detectar circunstancias tales como accidentes o delitos flagrantes.[751]

Este tipo de sistemas ofrece una amplia gama de posibilidades, siendo las más extendidas la de dar visibilidad a figuras y/o símbolos que no pueden ser captados por el ojo humano; mejorar la calidad de las imágenes, otorgándoles mayor nitidez; reconocer patrones; hacer mediciones; etc.

No obstante, no siempre las imágenes disponibles para las autoridades policiales y judiciales son de la mejor calidad. Y, en relación con ello, existen numerosas iniciativas y proyectos que tienen como objeto mejorar la calidad de estas para aumentar las posibilidades de éxito de su análisis. Entre otros, investigadores de Dartmouth College (Nuevo Hampshire, EEUU) están utilizando algoritmos que degradan sistemáticamente la calidad de las imágenes de alta

[751] Véase Rigano, 2019.

resolución y las comparan con las de baja resolución para reconocer mejor las imágenes y los vídeos de calidad. Así, por ejemplo, las imágenes claras de números y letras se degradan lentamente para emular imágenes de baja calidad, y dichas imágenes degradadas se catalogan como representaciones matemáticas que se pueden comparar con imágenes de matrículas de baja calidad para ayudar a identificar alguna en concreto.

Y es interesante ir más allá y poner de manifiesto que en los últimos tiempos se está investigando una una nueva técnica de "comprensión de la escena" (en inglés, "*scene understanding*") que tiene por objeto crear algoritmos con capacidad de desarrollar un texto que describa la relación existente entre personas, lugares y cosas presentes en una serie de imágenes con el fin de proporcionar un contexto. Así, por ejemplo, el texto podría ser "*La pistola está siendo desenfundada por una persona y descargada en el escaparate de una tienda*", y el objetivo no es otro que el de detectar objetos y actividades que ayuden a identificar delitos en curso (para llevar a cabo intervenciones en vivo), y desarrollar las investigaciones posteriores a los hechos. Y es que el análisis de una escena entre múltiples de ellas puede mostrar eventos potencialmente interesantes que deberán ser ulteriormente comprobados por las autoridades policiales.

En relación con ello, entre otros, un grupo de investigadores de la Universidad Central de Florida (EEUU), en asociación con el Departamento de Policía de Orlando, ha empleado fondos del National Institute of Justice (NIJ) para desarrollar algoritmos con el fin de identificar y relacionar "objetos" en videos (tales como personas, automóviles, armas y edificios), sin intervención humana.[752]

a.2) *Posibles utilidades en la instrucción de las causas*

a.2.1) Análisis (y eventual moderación) de contenido on line

Al igual que se ha puesto de manifiesto con anterioridad respecto de las herramientas que emplean técnicas de PLN, los sistemas de reconocimiento de imágenes pueden contar con un

[752] Véase Mubarak, 2018.

gran valor para rastrear y analizar contenido colgado en la red de Internet y, en su caso, detectar aquel que pueda resultar delictivo.

Hoy en día, las grandes empresas tecnológicas (Facebook, Youtube, etc…) ya han incorporado sistemas de IA que reconocen imágenes sensibles cuyo contenido puede resultar constitutivo de delito y bloquean de inmediato las cuentas que las comparten. Así, el 98% del contenido malicioso en Facebook está filtrado por algoritmos de *Machine Learning*[753] (los usuarios son los que denuncian el 2% restante); Google, propietario de YouTube, elimina el 80% de los videos inapropiados antes de que reciban cualquier visualización[754]; y Twitter informa de que pone el foco sobre unas diez cuentas por segundo.[755]

Ello, desde luego, es de enorme utilidad para parar la difusión, por ejemplo, de imágenes de pornografía infantil o de torturas, por ejemplo.

Así, en 2018 las empresas tecnológicas reportaron al gobierno de EEUU la cifra rércord de cuarenta y cinco millones de fotografías y videos de abusos sexuales a menores detectadas en la red (hace una década, el número reportado era de menos de un millón)[756], lo cual demuestra el gran valor que tienen las herramientas de IA para luchar contra este tipo de delitos.

Y es que, antes de que proliferara el uso de esta clase de herramientas de IA, las vías de detección de este tipo de material eran muy limitadas, habida cuenta de la escasa colaboración ciudadana (puesto que la mayoría de las personas que tienen acceso a este tipo de imágenes suele estar involucrada en los hechos delictivos) y de que los cuerpos policiales cuentan con recursos humanos muchas veces escasos y, si bien siempre han existido divisiones encargadas de rastrear la red y detectar posibles ilícitos penales, lo cierto es que su capacidad de detección era limitada, por lo que poder contar con sistemas que pueden analizar millones de imágenes en tiempo récord, es un adelanto muy significativo.

[753] Véase Facebook, 2019.
[754] Véase Comisión Europea, 2021.
[755] Véase Comisión Europea, 2021.
[756] Keller & Dance, 2019.

Además, este tipo de herramientas, generalmente combinadas con sistemas que emplean datos biométricos y que emplean técnicas de PNL, permiten rastrear y detectar la actividad de organizaciones o grupos criminales, incluidos terroristas.

En concreto, a modo de ejemplo, resulta interesante hacer especial mención al uso de las herramientas de IA empleadas por las distintas fuerzas policiales para detectar y prevenir (aunque también investigar) específicamente los delitos de tráfico sexual. En relación con ello, procede poner de manifiesto que los programas de *software* más potentes y extendidos en tal ámbito son fundamentalmente tres: Spotlight, Memex y Traffic Jam[757], siendo la mayoría de datos que emplean tales sistemas extraídos de anuncios *on line* de juguetes sexuales[758], siendo estimado que cada día se publican entre cuatrocientos mil y quinientos mil nuevos anuncios, críticas e hilos de discusión[759] relacionados con los mismos.

Las técnicas empleadas por los mencionados programas de *software* son muy completas y sofisticadas. Y es que "*Los anuncios se pueden analizar con técnicas de minería de datos basadas en pistas visuales, por ejemplo, en caso de que haya tatuajes presentes en varias imágenes; o señales textuales, incluyendo estilos específicos de redacción de anuncios, palabras clave, etc. para ayudar a diferenciar los anuncios sospechosos de los no sospechosos. Otro punto clave de la analítica avanzada impulsada por la IA es el de agrupar los anuncios en función de los auténticos propietarios y no de los pretendidos autores. La resolución de los casos implica que la IA descubra un conjunto de vínculos explícitos entre informaciones extraídas de diferentes fuentes. Como resultado, los datos tales como los números de teléfono y las direcciones de correo electrónico se pueden utilizar para encontrar patrones y conectar distintos anuncios que probablemente estén publicados por los mismos traficantes en diversas áreas geográficas.* (…) *Algunos software, incluidos Traffic Jam y XIX, también se basan en sistemas de análisis de imágenes. Tales herramientas pueden detectar imágenes o identificar fotografías similares que aparecen en anuncios aparentemente no conectados. Traffic Jam, asimismo, emplea un software de reconocimiento facial para identificar a las víctimas de tráfico sexual. Así, las fuerzas policiales pueden comparar la foto de una persona desaparecida con otras fotos que constan en la base de datos con el objetivo de identificar si esta está anunciada on-line. Finalmente, la analítica impulsada por IA*

[757] Véase Fox, 2017.
[758] Rizzi & Pera, 2020, pág. 75.
[759] *Idem*. Pág. 74.

también puede ayudar a identificar enlaces y nodos críticos dentro de una red y su estructura, siendo posible, posteriormente, exponer los patrones y tendencias más relevantes, tales como relaciones significativas, personas clave en el núcleo de la red y flujos entre nodos (por ejemplo, de bienes, finanzas o información)."[760]

No obstante, en muchas ocasiones los investigadores se topan con la escasa calidad de las imágenes, por lo que resulta muy difícil culminar con éxito las investigaciones, inlcuso empleando sistemas que mejoran la nitidez y demás atributos de estas.

Sin embargo, ya hay empresas explorando la posibilidad de preconstituir prueba con las imágenes provenientes del campo de batalla de Siria para presentarla ante la Corte Penal Internacional, donde, con todas las garantías, debería ser sometida a contradicción en el acto del plenario a los efectos de determinar la realidad material de los hechos y, en su caso, la existencia de errores y/o fallos en los sistemas de análisis de imágenes.

Así, en concreto la ONG estadounidense Benetech, a través de la iniciativa "*Connected Civil Society Initiative*", en colaboración con "*Syrian Civil Society Organizations*" (SCOs) y el Mecanismo Internacional Imparcial e Independiente (IIIM) de las Naciones Unidas para investigar atrocidades, está construyendo una plataforma digital que aplica técnicas de aprendizaje automático, de visión artificial y de análisis de metadatos para clasificar, identificar y analizar imágenes que podrían constituir pruebas de vulneraciones de derechos humanos y posibles crímenes de guerra.

Dicha plataforma facilita la coordinación entre la sociedad civil y ofrece tiempos récord de procesamiento de imágenes (dos o tres segundos por video); tiene capacidad para reconocer objetos tales como helicópteros, columnas de humo y tanques, así como para organizar y agrupar automáticamente los datos; detecta la cantidad de escenas que hay en un video y crea una huella digital para cada una con tal de aumentar la precisión de la agrupación por relación o coincidencia, etc, tal y como me expuso la abogada estadounidense Shabnam Mojtahedi, implicada en el proyecto.

[760] Rizzi & Pera, 2020, pág. 75.

Además, procede hacer referencia a la organización Syrian Archive, que tal y como dispone en su página web[761] tiene como objetivo apoyar a los investigadores, defensores de derechos humanos, reporteros y periodistas en sus esfuerzos por documentar las vulneraciones de los derechos humanos que tienen lugar en Siria y en todo el mundo mediante el desarrollo de nuevas herramientas de IA de código abierto, así como proporcionar una metodología transparente para recopilar, preservar y verificar e investigar documentación visual de áreas de conflicto. Y es que mediante tales tareas (basadas en el "*Electronic Discovery Reference Model*"[762] desarrollado en la Facultad de Derecho de la Univesidad de Duke, Carolina del Norte, EEUU), dicha organización tiene como objetivo preservar la información como una memoria digital y establecer una base de datos verificada de violaciones de derechos humanos que pueda emplearse como prueba para ser presentada ante la justicia, siendo este el procedimiento seguido:

763

Así, para la recopilación de datos se emplean fuentes fiables (presentaciones directas por parte de individuos y organizaciones o información públicamente disponible) y se asegura su almacenamiento en servidores, con copias de seguridad. Durante tal proceso se estandariza el formato de los datos (conservando el formato antiguo), que obtienen su *hash* y su sello o marca de tiempo (para garantizar que no se han modificado ni manipulado), y una vez que el contenido se ha conservado de forma segura, los metadatos se extraen del contenido visual, se analizan y se agregan automáticamente.

[761] Syrian Archive, s.f..
[762] Véase Electronic Discovery Reference Model, s.f..
[763] Syrian Archive, s.f..

Tales metadatos incluyen una descripción del objeto visual; la fuente del contenido; el enlace original donde se publicó por primera vez; el clima o tiempo (que puede ser útil para la geolocalización o la identificación de la hora); los idiomas específicos o dialectos regionales hablados (a través de técnicas de PLN); ropa o uniformes identificables; armas o municiones; dispositivo utilizado para grabar el metraje; y tipo de contenido multimedia, entre otros.

Posteriormente, se lleva a cabo la verificación, que consta de tres pasos: 1) verificar la fuente de quien subió el video o del editor; 2) verificar la ubicación donde se filmó el video; 3) verificar las fechas y horas en las que se grabó y subió el video.

Y, finalmente, una vez que el contenido ha sido procesado y verificado, se procede a su análisis y revisión para verificar su precisión, y si el contenido se considera preciso, se puede llegar a investigar y a considerar prueba digital.

b) Herramientas de lectura de matrículas

b.1) *Concepto*

Asimismo, es importante hacer especial mención de los sistemas de IA que tienen por finalidad la detección e identificación de matrículas -en inglés "*Automated Number Plate Recognition*" (ANPR)-, que en concreto emplean técnicas de OCR (Reconocimiento Óptico de Caracteres).

Tal tecnología, empleada por muchos cuerpos policiales de alrededor del mundo, permite, generalmente mediante el uso de cámaras instaladas en los coches patrulla, detectar aquellas matrículas que constan en una base de datos por pertenecer a un vehículo robado, por tener alguna incidencia administrativa, por haber estado su titular envuelto en la comisión de algún delito, etc.

Así, al igual que se ha puesto de manifiesto respecto de las demás herramientas que emplean técnicas análisis de imágenes, estas pueden contar con un gran valor para rastrear y detectar y, en su caso, analizar aquel contenido que pueda tener interés policial y/o judicial.

En relación con ello, procede hacer especial mención a la Decisión 2008/615/JAI del Consejo, de 23 de junio de 2008, sobre la profundización de la cooperación transfronteriza, en particular en materia de lucha contra el terrorismo y la delincuencia transfronteriza (conocida como Decisión Prüm), que en su artículo 12 hace referencia a la posibilidad de acceso por parte de los Estados miembros a las bases de datos nacionales en que consten matrículas de vehículos, lo cual incrementa exponencialmente el volumen de información disponible y establece:

"*1. Los Estados miembros permitirán que los puntos de contacto nacionales de los demás Estados miembros mencionados en el apartado 2, para los fines de la prevención y persecución de delitos y de la persecución de infracciones de otro tipo que sean competencia de los tribunales o de las fiscalías del Estado miembro que realice la consulta, y para el mantenimiento del orden público, tengan acceso a los siguientes datos contenidos en los registros nacionales de vehículos, con derecho a consultarlos de forma automatizada en casos concretos:*

a) datos de los propietarios o usuarios, y b) datos de los vehículos.

La consulta solo podrá efectuarse utilizando un número de bastidor completo o una matricula completa. La consulta deberá efectuarse con arreglo al Derecho interno del Estado miembro que la realice.

2. A efectos de las transmisiones de datos a que se refiere el apartado 1, cada Estado miembro designará un punto de contacto nacional para recibir solicitudes. Las competencias de los puntos de contacto nacionales se regirán por el Derecho interno que les sea aplicable. Los pormenores técnicos del procedimiento se regularán en las medidas de ejecución a que se refiere el artículo 33."

En concreto, y a modo de ejemplo, en Reino Unido el uso de cámaras de lectura automática de matrículas aumentó de dos mil en 2006 a nueve mil en 2017. Estas cámaras, según el informe anual (2016/2017) emitido al Parlamento de tal país por el Comisionado de Cámaras de Vigilancia, ya en ese momento escaneaban diariamente de veinticinco a cuarenta millones de matrículas de automóviles, y tales datos se almacenaban durante doce meses, lo que creó

una de las bases de datos no militares más grandes del Reino Unido con hasta veinte mil millones de registros.[764]

Como consecuencia de ello, no obstante, las voces críticas de las organizaciones *pro* derechos humanos no tardaron en llegar. Así, desde Big Brother Watch se alegó que no existía una base legal para el uso de tal tipo de herramientas y que, además, la retención de los datos de ubicación y su uso nunca habían sido acordadas por el Parlamento. En concreto, además, en su informe publicado en junio de 2018[765], dicha ONG puso de manifiesto que la jurisprudencia más reciente del Tribunal de Justicia de la UE[766] y del Tribunal Europeo de Derechos Humanos[767] también sugería que dicha recopilación y retención de datos de ubicación podía ser incompatible con el artículo 7 y el artículo 8 de la Carta de los Derechos Fundamentales de la UE o el artículo 8 del Convenio Europeo de Derechos Humanos.

En España, entre otras, en 2018 se instaló un sistema de lectura y detección automática de matrículas en los nuevos vehículos de la Guardia Urbana de Barcelona que, según afirma el Ayuntamiento de dicha localidad, localizó en seis meses unos dos mil doscientos vehículos sustraídos, con requerimiento policial o que habían sido precintados y, por ende, no podían circular.[768] A día de hoy, un total de doce vehículos está dotado de este tipo de sistemas.

Tal y como se expone por el antedicho Ayuntamiento, el detector de matrículas instalado "*es un sistema informático conectado a dos lectores en el puente de luces del vehículo, uno a cada lado, fijo y orientado hacia el suelo, que permite reconocer los caracteres alfanuméricos de las matrículas de los vehículos. De esta manera, un coche patrulla puede hacer la lectura de matrículas de los vehículos que se encuentran en los carriles de circulación como de estacionamiento de la derecha y de la izquierda de la calle. Los lectores son capaces de leer en unas 40 velocidades de fotogramas y al incorporar infrarrojos, pueden también leer las matriculas por la noche, a la sombra, en los túneles y en los aparcamientos*."[769]

[764] Véase Surveillance Camera Commissioner, 2018.
[765] Véase Big Brother Watch, 2018.
[766] Entre otros, casos C-203/15 *Tele2* Sverige AB v Post-och telestyrelsen; y C-698/15 SSHD v Tom Watson.
[767] Entre otros, Taylor-Sabori v United Kingdom; y Uzun v Germany.
[768] Ajuntament de Barcelona, s.f.
[769] Ajuntament de Barcelona, s.f.

Y es que los listados de matrículas de vehículos con interés policial se cargan y se actualizan de forma automática en el ordenador que el vehículo lleva a bordo, de modo que cuando las cámaras detectan una matrícula que consta en la base de datos, éste genera un aviso a través de una pantalla.[770]

Por su parte, es interesante hacer alusión a lo dispuesto en Resolución de 16 de julio de 2020, de la Subsecretaría de Estado de Interior, por la que se publica el Convenio entre el Centro para el Desarrollo Tecnológico Industrial, E.P.E., y el Ministerio del Interior, relativo a la contratación precomercial de servicios de I+D en materia de seguridad en el medio rural, que propone el uso del reconocimiento automático de matrículas como herramienta para el control de eventos multitudinarios. Así, en relación con ello hace referencia a la posibilidad de "*Identificación de personas con asuntos pendientes con la justicia en los puntos de control de acceso al evento*" y, al respecto dispone:

"*Se prevé que la solución innovadora pueda basarse, de manera general, en un sistema tecnológico innovador e inteligente formado por los siguientes componentes:*

• *Sistema de reconocimiento de matrículas. A ser instalado de manera previa a la entrada del aparcamiento del evento (por ejemplo, podría ser instalado en postes fijos o en vehículos aéreos no tripulados). Su propósito es proporcionar a los agentes que controlan los accesos al aparcamiento alertas para detener a vehículos con asuntos pendientes con la justicia*."

b.2) ***Posibles utilidades en la instrucción de las causas***

La posibilidad de detectar de forma automática, entre una multitud de vehículos, aquellas matrículas de interés para la instrucción de una causa, es sin duda un gran avance para los investigadores.

Y es que, imaginemos una causa por delito de tráfico de sustancias estupefacientes en que los principales sospechosos acuden a una nave industrial para cargar en un vehículo una ingente cantidad de cocaína. Pensemos, asimismo, que tales individuos están siendo sometidos a

[770] Ajuntament de Barcelona, s.f.

vigilancia de una patrulla policial que, al detectar tal movimiento, intenta proceder a su detención sin éxito, habida cuenta de la velocidad a la que dichos personajes consiguen huir con su turismo de alta cilindrada.

En tal caso, si no se contara con un sistema de IA de detección automática de matrículas en tiempo real, lo más probable sería que la patrulla actuante, por un lado, comunicara por radio al resto de agentes el número de matrícula del vehículo fugado por si se lo cruzaban y, por otro lado, la introdujera en el sistema policial para que saltara una alerta en caso de que fuera detectado en un control, etc. Ello, no obstante, implicaría tener que esperar a que algún agente diera con ellos, lo cual resulta complicado. No obstante, con un sistema de IA de detección de matrículas en tiempo real, las placas del turismo podrían ser detectadas tanto por las cámaras de la vía pública dotadas del mismo, como por los coches patrulla de la policía que sin duda saldrían "a la caza" o aquellos que simplemente se dedicaran a patrullar días más tarde, lo cual multiplicaría las posibilidades de éxito.

c) Herramientas de detección de documentos falsos

c.1) *Concepto*

Finalmente, resulta interesante hacer referencia a las herramientas de IA que, a partir del análisis de los elementos de un documento, pueden determinar si este está o no falsificado.

En relación con ello, por ejemplo, hoy en día el Cuerpo Nacional de Policía cuenta con el sistema IFADO ("*Intranet False and Authentic Documents Online*") que contiene una base de datos con la información más importante sobre los documentos oficiales (de viaje y de identidad) para su intercambio entre expertos, con el fin de determinar si estos son auténticos o falsos. Y es que en tal sistema están introducidas las características específicas de los mencionados documentos y cuando los agentes dudan de la autenticidad de uno de ellos lo que hacen es compararlo con las imágenes indubitadas contenidas en la base de datos de tal sistema.

No obstante, por el momento no se emplean técnicas de análisis de imágenes automatizado, sino que la comparación la hacen a mano los expertos humanos, por lo que la eficacia de tales operaciones, desde luego, podría resultar enormemente incrementada en caso de emplear técnicas de IA, habida cuenta de que estas reducirían los tiempos de resolución de los casos y podrían aumentar la precisión si se contrara con sistemas potentes cuyos resultados, no obstante, sin duda, podrían ser verificados por un experto humano.

c.2) ***Posibles utilidades en la instrucción de las causas***

Y es que imaginemos que la policía procede a detener a un ciudadano que entrega un pasaporte falso, aparentemente auténtico. Hoy en día, ello, a los ojos de los agentes de policía encargados de la detención, en caso de tratarse de una buena falsificación, podría pasar por alto. No obstante, si los cuerpos policiales, por ejemplo, contaran con sistemas de IA de detección automática de documentos falsos, sería deseable que, por protocolo, cada documento presentado por un ciudadano fuera sometido de forma inmediata a control, de forma que se incrementaría exponencialmente la detección de casos de falsedad documental y restaría presión a los agentes encargados de gestionar la detención.

C.3. ***Regulación española y europea***

Respecto de ello y, al igual que he puesto de manifiesto en la Sección anterior, en aras de evitar caer en reiteraciones, habida cuenta de que el uso de los sistemas analizados puede implicar, con toda probabilidad, el tratamiento de datos personales, procede remitirse a lo dispuesto, con carácter general, respecto de la regulación de las herramientas de IA que emplean datos biométricos (habida cuenta de que estos tienen la consideración de datos de carácter personal). Y es que todo lo que les resulta aplicable es extensible a las herramientas que ahora nos ocupan, con la salvedad de lo dispuesto de forma específica para tal clase de sistemas.[771]

Así, procede hacer mención, de forma principal, a lo dispuesto en la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales,

[771] Véanse págs. 416-440.

y a lo previsto en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión publicada por la Comisión Europea el 21 de abril de 2021, si bien tal y como ya se ha venido poniendo de manifiesto ello no es más que una mera propuesta, como su propio nombre indica, y por ende, debe tomarse con las oportunas cautelas.

C.4. ***Riesgos jurídicos generales***

Con carácter general, las herramientas de IA analizadas se enfrentan a los mismos potenciales peligros que las analizadas con anterioridad, en los términos ya expuestos, que se dan por reproducidos.

No obstante, de forma más concreta, al hablar del uso de tal clase de herramientas debe atenderse, en primer lugar, a su posible falta de precisión y a su potencial discriminatorio.

Y es que, por un lado, estas emplean sistemas informáticos cuya potencia y nivel de sofisticación puede influir en el resultado que arrojan, y por desgracia no siempre se cuenta con la tecnología más puntera (habida cuenta de que esta es muy cara); y, por otro lado, estas se nutren de datos y estos pueden revestir una calidad cuestionable (por ejemplo, es muy posible que los cuerpos policiales europeos dispongan de mejores bases de datos con la información más importante sobre los documentos oficiales de los países occidentales).

En segundo lugar, debe prestarse atención a la posible vulneración del derecho a la privacidad y a la protección de datos personales, habida cuenta de que tanto los lectores de matrículas, que realizan un tratamiento de datos personales en tanto en cuanto estas son una numeración capaz de identificar a una persona física (el titular del vehículo), como los detectores de documentos falsos, que realizan un tratamiento de datos personales en cuanto analizan los datos que constan en los documentos oficiales, están sujetos a la legislación en materia de protección de datos personales.

En tercer lugar, debe ponerse el foco en las posibles brechas de seguridad que se pueden generar en relación con las mencionadas herramientas de IA, habida cuenta de que contienen

información muy sensible y las consecuencias de su vulnerabilidad podrían resultar fatales. Así, por ejemplo, un cibertaque a un sistema de análisis de imágenes que tuviera como finalidad detectar posibles crímenes de guerra o averiguar las circunstancias en que estos fueran cometidos, por ejemplo, podría poner en jaque el buen curso de una investigación judicial; y, una filtración de los datos de un sistema de lectura de matrículas empleado por la policía, por ejemplo, podría asimismo suponer un grave perjuicio para los titulares de los vehículos que constan registrados en el mismo, ya que podrían ver publicados en la red circunstancias relativas a su relación con la justicia (en ocasiones, de forma errónea).

Y, finalmente, debe ponerse atención a la posible falta de transparencia que este tipo de sistemas puede entrañar, puesto que, siendo que su uso puede conllevar la toma de decisiones policiales y/o judiciales que afecten a los derechos fundamentales de las personas, es indispensable que tanto estas como las autoridades que las emplean puedan tener conocimiento real de cómo funcionan y con base en qué arrojan sus resultados en aras de poder rebatirlos y detectar, en su caso, posibles deficiencias o fallos.

D- Herramientas que emplean otras tecnologías

D.1. *Concepto*

A modo conceptual, este tipo de herramientas hace referencia a todos aquellos sistemas de IA que no emplean ninguna de las tecnologías ya expuestas con anterioridad, por lo que opera como "cajón de sastre" en tal sentido.

D.2. *Subclases*

a) Herramientas de detección de estafas o fraudes digitales

a.1) *Concepto*

Una gran parte de las empresas que ofrecen servicios *on line* y, especialmente, las plataformas de pagos, emplean herramientas de IA para detectar y combatir comportamientos fraudulentos, a saber, principalmente, la creación de cuentas de usuario falsas y la realización

de pagos ficticios, lo cual resulta de enorme utilidad, ya que sin dicha tecnología, ello sería un problema prácticamente imposible de resolver.

Este tipo de herramientas funciona recopilando y analizando ingentes conjuntos de datos y entrenando continuamente a sus algoritmos para predecir y reconocer patrones anómalos. Entre los datos analizados se hallan: el correo electrónico, el número de teléfono, los metadatos del dispositivo desde el que se realiza la operación, la dirección IP, los perfiles de redes sociales (para evaluar la legitimidad de direcciones de correo electrónico y números de teléfono, por ejemplo), etc y con base en ellos se realizan un análisis de comportamiento del usuario, combinando el uso de potentes sistemas de *Machine Learning* y la inteligencia humana para asignar puntuaciones de riesgo a las acciones de los usuarios, lo que hace saltar la alarma cuando se advierte de que algo es potencialmente anómalo o directamente aprueba o rechaza de forma inmediata la transacción que se pretende llevar a cabo.

En relación con ello, es interesante hacer especial referencia al caso de la plataforma de pagos PayPal. Y es que tal y como su Director de Riesgos, Hui Wang, manifestó, ya en 2015, la compañía utiliza tres tipos de algoritmos de *Machine Learning* para la gestión de riesgos: lineal, de red neuronal y *Deep Learning*, habiendo demostrado su experiencia que el enfoque más eficaz es el de hacer uso de los tres a la vez.[772]

En relación con ello, el mencionado directivo puso de manifiesto: "*Nos tomamos la confianza muy en serio. Es nuestra marca. Tenemos que decidir en un par de cientos de milisegundos si estamos ante una buena persona (en cuyo caso le daremos la mejor y la más rápida experiencia), o si estamos ante un tipo potencialmente malo y tenemos que tomar decisiones (...) Determinar rápidamente a los clientes confiables y ponerlos en el carril rápido para una transacción es un objetivo clave*".[773] Y añadió: "*Los algoritmos más sofisticados se aplican a los clientes que pueden ser problemáticos, lo que ralentiza un poco el sistema a medida que adquiere más datos para realizar un análisis en profundidad*."[774]

a.2) ***Posible utilidad en la instrucción de las causas***

[772] Knorr, 2015.
[773] Knorr, 2015.
[774] *Idem.*

Hoy en día los juzgados están llenos de asuntos de presuntas estafas cometidas en Internet que en la mayoría de casos conllevan un arduo trabajo policial para dar con los autores de los hechos y un gran despliegue de medios judiciales, siendo que en muchas ocasiones son necesarias comisiones rogatorias que tardan en llegar, autorizaciones judiciales de análisis de teléfonos, cuentas bancarias etc, que, por desgracia, generalmente acaban cayendo en saco roto dada la dificultad de investigar delitos que no solo traspasan nuestras fronteras sino que además suelen ser cometidos desde distintos puntos del planeta por verdaderos profesionales.

Así, la existencia de sistemas que ayuden a las propias compañías que operan en Internet a detectar posibles fraudes de forma temprana, no solo ayuda a prevenir la comisión delictiva (siendo que por lo general los delitos que trataran de cometerse quedarían en meras tentativas) sino que puede ayudar asimismo a su investigación.

Y es que imaginemos que un delincuente sustrae una tarjeta de crédito a un turista extranjero y al cabo de los días trata de hacer con ella una compra *on line*. En caso de no existir sistemas de IA que analizaran la operación antes de que esta fuera completada, el mencionado sujeto podría efectuar la adquisición con mayor facilidad; en cambio, en caso de que la plataforma de pagos empleara tal clase de herramientas, en el mismo momento en que el delincuente introdujera la numeración de la tarjeta bancaria en la mencionada plataforma, esta detectaría que se trataba de una tarjeta extranjera, que la compra se estaba intentando llevar a cabo desde una dirección IP española, que los datos del usuario eran españoles y los del titular de la tarjeta extranjeros etc y, por ende, como mínimo, haría saltar una alerta que implicaría la exigencia de un plus de verificación para autorizar la operación.

Ello, desde luego, en caso de que no llegara a superarse el control de seguridad, podría ser reportado a las autoridades para que estas procedieran a iniciar la correspondiente investigación, partiendo ya de unos datos altamente valiosos para dar con la identidad de los presuntos autores de los hechos.

b) Herramientas de detección de disparos

b.1) *Concepto*

El uso de armas es uno de los problemas más extendidos en el mundo. Y es que, si bien estas en algunas ocasiones son empleadas por de las fuerzas militares o del orden con fines legítimos, en muchas otras son utilizadas por individuos con finalidades ilegítimas.

En concreto, en Occidente, uno de los países que cuenta con mayor número de incidentes por uso indiscriminado de armas por parte de sus ciudadanos, habida cuenta de la flexible regulación existente en tal país, es Estados Unidos. Como consecuencia de ello, es allí donde la tecnología que emplea IA para detectar en tiempo real la existencia de disparos o tiroteos con la finalidad de evitar situaciones de riesgo, está más desarrollada.

En concreto, procede hacer especial mención al denominado sistema "*ShotSpotter Gunshot Detection*". Este, desarrollado por la compañía ShotSpotter, que en su página web pone de manifiesto que el uso de armas ha causado la muerte de más de sesenta y cinco mil personas y ha lesionado a varios cientos de miles en Estados Unidos en los últimos cinco años (y lo más llamativo es que el 80% de los incidentes con disparos nunca fueron reportados o denunciados ante la policía), tiene como objetivo detectar las posibles situaciones de riesgo generadas por disparos, generar una eficaz y rápida respuesta policial y recopilar indicios y pruebas que luego puedan ser presentadas ante las autoridades fiscales y judiciales.

Dicho sistema, utilizado fundamentalmente por los cuerpos policiales de EEUU, emplea diversos sensores acústicos ubicados en diversos puntos del área geográfica en que va a usarse, que detectan en tiempo real, con IA, el sonido de un disparo y la triangulación permite determinar la ubicación precisa (en función del tiempo que tarda el sonido en llegar a cada uno de los sensores), y manda la información a una centralita. En dicha centralita, un grupo de expertos de la compañía verifica en tiempo récord si efectivamente se trata de un disparo y, en caso afirmativo, lo reporta a la policía.[775]

Ello, desde luego, tiene una enorme utilidad para mitigar los efectos de un posible delito, por la rapidez con la que se pone en conocimiento de la policía. Y, asimismo, tiene gran utilidad

[775] Véase ShotSpotter, s.f..

para posteriormente investigar los hechos, ya que se cuenta con información precisa relativa a la existencia de los disparos y a la hora y el lugar de los mismos.

Dicha tecnología, por los grandes beneficios que puede reportar, está en constante evolución. Así, por ejemplo, en un proyecto financiado por el National Institute of Justice (NIJ) norteamericano, los científicos de la compañía Cadre Research Labs, LLC desarrollaron un sistema de IA basado en un modelo matemático que contenía algoritmos entrenados para ir más allá de la mera detección de un disparo y poder diferenciar entre tipos de disparo, determinar la cantidad de armas de fuego presentes, asignar disparos a cada una de las armas de fuego participantes y estimar probabilidades de la clase y el calibre de las mismas, lo cual podría ayudar enormemente a las autoridades policiales, fiscales y judiciales en la tarea de investigación criminal.[776]

b.2) ***Posibles utilidades en la instrucción de las causas***

Es evidente que toda herramienta que suponga acortar los tiempos de respuesta ante posibles delitos y aumentar la eficacia de la actuación policial (y, por ende, posteriormente, fiscal y judicial), es beneficiosa para los ciudadanos.

En este caso, desde luego, el hecho de que cualquier disparo sea percibido en tiempo real y reportado a las autoridades policiales en tiempo record, supone un avance sin precedentes en la lucha contra los delitos cometidos con armas.

Y es que, imaginemos que se produce un tiroteo por parte de diversos terroristas que tienen intención de seguir disparando a la gente de forma indiscriminada. Pensemos que, no obstante, el primer disparo es captado por los sensores acústicos y es geolocalizado, por lo que de forma inmediata hace saltar una alarma en la centralita del sistema, donde se verifica no solo que se trata de un disparo, sino que además, al ir llegando los distintos sonidos, se determina cuántas armas hay, qué clase de armas son, qué tipo de disparos se están llevando a cabo etc, información que es reportada a la policía y que es valiosísima para que estos tomen la decisión de respuesta más adecuada. Y es que pensemos que la información que da el sistema es que se trata de 4 rifles Kalashnikov, que están disparando de forma indiscriminada, y que van

[776] Véase National Institute of Justice NIJ., EEUU, 2016.

moviéndose a lo largo de X calle, lo cual muy probablemente hará pensar enseguida a la policía que se trata de un delito terrorista y, por ende, movilizarán ya de entrada todos los recursos que ello requiere para dar una respuesta contundente y efectiva.

Además, si se trata de un sistema de IA potente, podría incluso determinar qué arma ha disparado a cada persona, lo cual a efectos de investigación penal es valiosísimo.

En caso de no contar con tal sistema, no obstante, si bien está claro que la policía hubiera sido igualmente alertada, lo más seguro es que esta hubiera recibido información más confusa y hubiera tenido un margen de reacción inferior, lo cual hubiera entorpecido no solo la evitación de más daño, sino también la ulterior investigación judicial.

D.3. *Regulación española y europea*

Respecto de ello y, del mismo modo que he manifiestado en la Sección anterior, en aras de evitar repeticiones, habida cuenta de que el uso de los sistemas analizados puede implicar el tratamiento de datos personales (no en el caso de las herramientas de detección de disparos pero sí en las de detección de estafas y fraudes digitales, por ejemplo), procede remitirse a lo dispuesto, con carácter general, respecto de la regulación de las herramientas de IA que emplean datos biométricos (habida cuenta de que estos tienen la consideración de datos de carácter personal). Y es que todo lo que les resulta aplicable es extensible a las herramientas que ahora nos ocupan, con la salvedad de lo dispuesto de forma específica para tal clase de sistemas. [777]

Así, procede hacer mención, de forma principal, a lo dispuesto en la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales y a lo previsto en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión publicada por la Comisión Europea el 21 de abril de 2021, si bien tal y como ya se ha venido poniendo de

[777] Véanse págs. 416-440.

manifiesto ello no es más que una mera propuesta, como su propio nombre indica, y por ende, debe tomarse con las oportunas cautelas.

D.4. *Riesgos jurídicos generales*

Con carácter general, al igual que se ha dispuesto respecto de las herramientas de IA que emplean técnicas de Visión Artificial o *Computer Vision*, las herramientas de IA analizadas se enfrentan a los mismos potenciales peligros que las analizadas con anterioridad, en los términos ya expuestos, que se dan por reproducidos.

No obstante, de forma más concreta, al hablar del uso de tal clase de herramientas debe atenderse, en primer lugar, a su posible falta de precisión y a su potencial discriminatorio

Y es que, por un lado, estas dependen de la potencia y del nivel de sofisticación de los *software* que emplean, cuya calidad lamentablemente no siempre es la mejor habida cuenta de sus elevados precios. Ello, no obstante, debería estar siempre sujeto a estrictos controles, puesto que sin duda no puede resultar legítimo emplear IA "a cualquier precio".

Y, por otro lado, tales herramientas, pueden implicar vulneraciones del derecho a la igualdad y a la no discriminación por el empleo de datos que no son lo suficientemente representativos o por su uso indiscriminado.

En relación con ello, especialmente interesante es hacer alusión a las voces críticas que han surgido respecto de sistemas de detección de disparos tales como ShotSpotter. Y es que los sectores más conservadores defienden que este tipo de herramientas presenta una amenaza real para los cidudanos de aquellas ciudades en las que están instalados, habida cuenta de que, por un lado, con frecuencia existen más sensores acústicos en zonas con residentes negros y latinos, lo que podría resultar discriminatorio; y, por otro lado, la precisión todavía deja bastante que desear, existiendo varios errores que hacen que los agentes acudan a determinados puntos e interpreten que cualquier persona que se halle cerca sea percibida como un posible delincuente de forma errónea. Y es que, en concreto, un estudio realizado por el MacArthur Justice Center de la Northwestern Pritzker School of Law (Chicago, Illinois, EEUU) que analizó las actuaciones de ShotSpotter en la ciudad de Chicago desde julio de

2019 hasta abril de 2021, concluyó que el 89% de los incidentes reportados a la policía no revelaron la existencia de ningún delito relacionado con armas y el 86% no dio lugar a ningún atestado.[778]

Ello llevó a la presentación ante los Tribunales, en abril de 2021, de un escrito de *amici curiae*[779] por parte de una coalición de organizaciones de Chicago preocupadas por el impacto del uso de ShotSpotter en dicha ciudad (especialmente en barrios donde residen personas de raza negra): Brighton Park Neighborhood Council, Lucy Parsons Labs y Organized Communities Against Deportations (OCAD), que ponía en cuestión la validez científica de los informes emitidos por el sistema ShotSpotter, que los fiscales estaban empezando a utilizar como prueba en los procesos penales.

En mi opinión, no obstante, a la vista de lo expuesto, un uso del sistema de modo uniforme por las distintas áreas del territorio de referencia y una mejora de su calidad, sin duda podrían hacerlo apto para los fines perseguidos, que son absolutamente legítimos y beneficiosos para la sociedad.

En segundo lugar, debe prestarse atención a la posible vulneración del derecho a la privacidad y a la protección de datos personales, especialmente en relación con el uso de sistemas de detección de estafas y fraudes, habida cuenta de que realizan tratamientos de datos personales y, por ende, están sujetos a la legislación en materia de protección de datos personales.

En tercer lugar, debe ponerse el foco en las posibles brechas de seguridad que se pueden generar en relación con las mencionadas herramientas de IA, habida cuenta de que un ataque o fallo en su funcionamiento o en su contenido puede resultar altamente perjudicial.

Y es que, por ejemplo, un cibertaque a un sistema de detección de disparos podría frustrar una intervención policial que, en caso de haberse llevado a cabo, hubiera evitado la muerte de algún ciudadano; y, un ataque a un sistema de detección de estafas y fraudes, sin duda podría

[778] Véase MacArthur Justice Center, 2021.
[779] Véase Brighton Park Neighborhood Council, Lucy Parsons Labs & Organized Communities against Deportations, 2021.

o bien dejar al descubierto miles de datos cruzados de personas anónimas o bien no detectar la comisión de ilícitos penales.

Y, finalmente, debe ponerse atención a la posible falta de transparencia que este tipo de sistemas puede entrañar, puesto que, al igual que se ha puesto de manifiesto respecto de las herramientas que emplean técnicas de Visión Artificial o *Computer Vision*, siendo que su uso puede conllevar la toma de decisiones policiales y/o judiciales que afecten a los derechos fundamentales de las personas, resulta innegociable que sean trazables y transparentes con la finalidad de que tanto los interesados como las autoridades que las emplean puedan tener acceso al contenido y la forma de funcionar de las mismas, a los efectos de discutir sus resultados o detectar posibles fallos.

-El fenómeno de las vigilancias masivas ("mass surveillance")

Una vez expuesto todo lo anterior, procede hacer referencia a un polémico fenómeno que ha ido cobrando protagonismo en los últimos años, en la era del *big data* y que está estrechamente vinculado con los sistemas de IA que emplean datos biométricos: las vigilancias masivas ("*mass surveillance*").

Como cuestión previa, para entender mejor a qué se hace referencia cuando se habla de vigilancias masivas, debe definirse el término abreviado *big data*.

En relación con ello, tal y como la Casa Blanca puso de manifiesto en un informe emitido en 2014: "*Existen muchas definiciones de "big data" que pueden diferir dependiendo de si usted es un informático, un analista financiero o un empresario que presenta una idea a un capitalista de riesgo. La mayoría de las definiciones reflejan la creciente capacidad tecnológica para capturar, agregar y procesar un volumen, una velocidad y una variedad de datos cada vez mayores*."[780]

Así, a grandes rasgos, se entiende que la expresión *big data* alude a "*la captación y el análisis de grandes conjuntos de datos con el objetivo de revelar informaciones o patrones ocultos*."[781]

[780] Gobierno de Estados Unidos, 2014, pág. 2.
[781] Cohen, 2013, págs. 1920-1921.

O, en otras palabras "*grandes conjuntos de datos que pueden ser ordenados por potentes computadoras para visualizar conexiones o correlaciones inesperadas.*"[782]

Y es que justamente el *big data* es la base necesaria para llevar a cabo las vigilancias masivas a las que hago referencia, habida cuenta de que estas no son más que actuaciones de control colectivo y generalizado de los distintos aspectos de la vida de los ciudadanos, para lo que sin duda se requieren ingentes cantidades de datos y potentes tecnologías que los analicen y establezcan relaciones entre todos ellos.

Respecto de ello, de forma muy gráfica, el Profesor de Derecho de la American University (Washington DC, EEUU) Andrew Ferguson, en su libro "*The Rise of Big Data Policing*", afirma: "*Estás siendo observado. Vigilado. Seguido. Dirigido. Cada búsqueda en Internet es registrada. Saben a qué velocidad conduces, cuáles son tus cereales preferidos, tu talla de vestido. Conocen tu situación financiera, todos tus trabajos anteriores, tu límite de crédito. Conocen tus preocupaciones de salud, preferencias de lectura y patrones de votación política. También conocen tus secretos. Te han estado observando durante años. En realidad, estás en un estado de vigilancia. Los observadores te conocen por los datos que dejas atrás. Pero no eres solo tú. Estos observadores también conocen a su familia, amigos, vecinos, colegas, clubes y asociados.*"[783]

Y es que ello, nos guste o no, es una realidad. En la actualidad, no obstante, la mayor parte de los beneficiarios de tales datos son compañías con ánimo de lucro, si bien cada vez son más las Administraciones Públicas interesadas en hacer acopio de tal información, especialmente con el fin de velar por la seguridad de sus ciudadanos (o, al menos, eso es lo que formalmente se afirma).

Hoy en día es especialmente fácil y viable llevar a cabo un seguimiento masivo y un control de los ciudadanos, no solo por las tecnologías que las empresas y los gobiernos tienen a su alcance para monitorizar a la población (cámaras de videovigilancia de identificación biométrica, cámaras de identificación automática de matrículas de los vehículos, etc) sino por

[782] Berman, 2013, pág. 6.
[783] Ferguson, 2017, pág. 7.

el rastro que los propios ciudadanos van dejando, especialmente con el uso de Internet y las nuevas tecnologías (compras *on line*, GPS, *smartphones*, elementos domóticos, etc).

En relación con ello procede poner de manifiesto que, a pesar de los enormes beneficios que el uso de tales prácticas de vigilancia masiva podría aportar en el ámbito de la investigación criminal (no solo por el brutal incremento de información que generaría para las autoridades sino también por el gran impacto que estas tendrían en la prevención tanto general como especial), lo cierto es que los riesgos que lleva implícitos para los derechos fundamentales de los ciudadanos son enormes, incluso mayor a las eventuales utilidades. Y así lo ha considerado, entre otras, Amnistía Internacional, que entiende que las vigilancias masivas van en contra de lo dispuesto en el artículo 12 de la Declaración Universal de Derechos Humanos, y establece que: "*Nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra o a su reputación. toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques*."[784]; y la Comisión Europea, que en su Propuesta de Reglamento de IA prohíbe con carácter general, tal y como ya se ha expuesto con anterioridad, por ejemplo, los sistemas de videovigilancia biométrica remota en tiempo real salvo para casos muy específicos (artículo 5.1.d) y, asimismo, veta una de las prácticas que pueden potencialmente resultar más perjudiciales para los derechos de los ciudadanos en los casos de vigilancias masivas: las herramientas de IA de evaluación o clasificación de la confiabilidad de los ciudadanos aportándoles una puntuación social con el fin de dispensarles un trato perjudicial o desfavorable en contextos sociales no relacionados con aquellos en los que se generaron o recopilaron originalmente los datos, y/o dispensarles un trato perjudicial o desfavorable de forma injustificada o desproporcionada con respecto a su comportamiento social o su gravedad (artículo 5.1.c).

Por un lado, desde luego la opción de llevar a cabo vigilancias masivas con fines de seguridad es altamente tentadora. Y es que, imaginemos que llega a la policía una *notitia criminis* y esta, en vez de aplicar los métodos de investigación tradicionales (en muchas ocasiones lentos, costosos y poco eficaces, principalmente por la falta de recursos materiales y personales que acecha a la mayor parte de cuerpos policiales) simplemente tiene que recurrir a las cámaras con tecnología de reconocimiento facial instaladas en el lugar de los hechos, identificar al

[784] Véase Amnistía Internacional, s.f.

sospechoso de la comisión del delito y a partir de ahí rastrear todos sus movimientos anteriores y posteriores a la perpetración del mismo, pudiendo así de forma rápida y sencilla identificar otros implicados en los hechos, hallar el cuerpo del delito, detectar su participación en otros delitos, etc.

Así, pensemos en un caso real que lleva más de una década sin resolver, el de Marta del Castillo. Imaginemos que, en el momento en que la policía recibió la noticia de su desaparición, esta hubiera podido emplear un sistema de IA que en cuestión de minutos hubira rastreado todos sus últimos movimientos gracias al cruce de información de la geolocalización de su *smartphone*, de las compras efectuadas con su tarjeta bancaria, de las imágenes captadas por las cámaras de videovigilancia en tiempo real de la vía pública, etc y la hubiera ubicado en el domicilio de Miguel Carcaño, principal sospechoso de los hechos, de quien también hubiera también podido seguir el rastro y determinar hasta qué hora estuvo en su vivienda, dónde fue al salir de allí, si tuvo contacto con una o varias personas la noche de los hechos (y, en su caso, las identidades de estas), etc. Y sobre esa base, imaginemos que también la policía puede rastrear los movimientos anteriores y posteriores de las personas que tuvieron contacto con el principal sospechoso de la desaparición. Ello, sin duda, desde el punto de vista de la investigación criminal, sería un auténtico lujo.

No obstante, hay que tener en cuenta que lo expuesto supone una captación masiva de los datos de los ciudadanos, lo que podría resultar infinitamente peligroso en caso de hacerse de forma indiscriminada e injustificada, habida cuenta de que ello puede implicar la ausencia de libertad total de la población.

En relación con esto último, procede prestar especial atención a lo que está ocurriendo en China.

Por un lado, en Guiyang (capital de la región de Guizhou, al suroeste de China) gracias a un vasto y sofisticado sistema de cámaras dotadas de tecnología de reconocimiento facial en tiempo real, la policía supuestamente puede localizar e identificar a cualquier persona que muestre su rostro en público en cuestión de minutos, así como rastrear dónde ha estado durante la última semana y, además, la policía puede relacionar el rostro de cualquier ciudadano con

su vehículo, sus familiares y personas con las que está en contacto o se encuentra de modo frecuente.[785]

Por otro lado, lo que ocurre en la región china de Sinkiang, va mucho más allá.

Como es públicamente conocido, en dicha región existe un conflicto activo entre el gobierno central y miembros de grupos separatistas de la etnia uigur -dirigidos principalmente por organizaciones islamistas turcas- que ya han causado varios incidentes (el más grave, el atentado que se produjo en 2014 en el mercado de Urumchi, la capital de la región). En relación con ello, desde finales del 2016 el gobierno chino, bajo la justificación de la campaña antiterrorista denominada "*Strike Hard Campaign against Violent Terrorism*", ha sometido en Sinkiang a trece millones de uigures y otros musulmanes turcos a vigilancias masivas que las autoridades califican como necesarias para combatir y prevenir el terrorismo, transmitiendo a la población que gracias a sus sofisticados sistemas de vigilancia la región permanece a salvo, ya que los potenciales terroristas son detectados de forma rápida y con alta precisión.

No obstante, tal y como denuncia la ONG Human Rights Watch en su informe "*China's Algorithms of Repression*"[786], dichas vigilancias han conllevado detenciones arbitrarias, adoctrinamiento político forzoso, restricciones de movimiento y represión religiosa, lo cual resulta absolutamente inaceptable. Así, en el seno de dicha campaña, según el antedicho informe las autoridades de Sinkiang han recopilado datos biométricos, incluyendo muestras de ADN, huellas dactilares, escáneres de iris e información sobre el grupo sanguíneo de todos los habitantes de la región de entre 12 y 65 años y, asimismo, han requerido una muestra de voz a los ciudadanos que solicitaban un pasaporte.

Tal y como se desprende de dicho informe, gracias a la utilización de técnicas de ingeniería inversa[787] se ha descubierto que la policía de Sinkiang utiliza una aplicación para establecer comunicación con el denominado "*Integrated Joint Operations Platform*" (IJOP), uno de los

[785] Véase BBC News, 2017.
[786] Human Rights Watch, 2019.
[787] La ingeniería inversa o retroingeniería es un proceso realizado con el objetivo de averiguar, a partir de un producto y mediante el razonamiento abductivo, cuál es su contenido, cómo fue su proceso de fabricación, etc.

principales sistemas empleados por las autoridades chinas para llevar a cabo vigilancias masivas en tal región. Y es que la IJOP *app* es empleada para, por un lado, recopilar información personal de los ciudadanos que luego es introducida en el sistema central y asociada al número de documento de identidad de cada persona; y, por otro lado, reportar actividades o circunstancias consideradas sospechosas y promover investigaciones de personas etiquetadas o calificadas como problemáticas por tal sistema. Por su parte, la tecnología de reconocimiento facial empleada por las cámaras de videoviglancia ubicadas en la vía pública, busca exclusivamente a los uigures en función de su apariencia (rasgos faciales) y registra todos sus movimientos para su ulterior búsqueda y revisión por parte de las autoridades.

Así, según se pone de manifiesto en tal informe, las vigilancias masivas realizadas en Sinkiang no tienen como prioridad la lucha antiterrorista. Y es que, entre otras cuestiones, las autoridades no se limitan a recopilar datos de los integrantes de los potenciales grupos terroristas, sino que captan datos de todos los habitantes de la región, rastrean el movimiento de los ciudadanos monitorizando la trayectoria y la localización de sus teléfonos móviles, documentos de identidad y vehículos, su consumo de electricidad y el uso de las gasolineras, entre otros, de modo que en el momento en que el sistema IJOP detecta alguna desviación de lo que considera "normal" (por ejemplo, el uso de un teléfono no registrado como propio o de alguien cercano, un aumento injustificado de consumo eléctrico, el uso habitual de puertas distintas a la principal para entrar o salir de la vivienda, un cambio de zona de residencia a otra distinta de aquella en que se estaba registrado sin autorización policial, el uso de herramientas de *network* tales como Whatsapp y Viber, etc), reporta tales micro indicios a las autoridades y provoca una investigación.

En relación con ello, se establecen treinta y seis tipos de personas a las que debe prestarse especial atención en función del nivel de amenaza que las autoridades perciben -determinado por los factores programados en el sistema IJOP- y, según el grupo al que pertenezcan, la libertad de los ciudadanos se restringe en un grado u otro (algunos son ingresados en prisiones de la región o en campos de educación política, otros quedan en arresto domiciliario, otros reciben la prohibición de acceder a lugares públicos o incluso de abandonar China sin permiso, etc.)

Es interesante poner de manifiesto, además, que la mencionada IJOP *app* es una herramienta multifuncional, puesto que tal y como afirma el mencionado estudio, no solo permite recopilar información y promover investigaciones policiales, sino que además, posibilita establecer comunicación entre los oficiales a través de la herramienta AcroPhone; planear las rutas más rápidas a través de un sistema de GPS; buscar toda la información disponible sobre una persona, un vehículo, etc al instante; detectar mediante tecnología de reconocimiento facial si la fotografía de un documento de identidad cuadra con la cara de una persona; comparar fotografías de documentos distintos; etc.

Según el antedicho informe, el apartado más interesante de la *app* es el denominado "misiones de investigación". Y es que las instrucciones de dichas misiones son enviadas directamente a través del sistema central IJOP a los oficiales, requiriéndoles para investigar a ciertos individuos, vehículos o eventos, o para remitir *feedback*. No obstante, también se permite a los oficiales presentar informes *motu proprio* sobre personas, objetos o vehículos que consideren sospechosos, pudiendo añadir incluso fotografías o grabaciones de voz.

Con el objetivo de que el gobierno central chino pueda controlar que los oficiales emplean el sistema IJOP, la aplicación les puntúa en función del uso que hagan de la misma. Y es que según el informe analizado, los oficiales del gobierno están sometidos a una enorme presión para llevar a cabo la denominada "*Strike Hard Campaign against Violent Terrorism*", ya que se señala y detiene a los oficiales que se sospecha que son desleales al régimen.

El estudio analizado, no obstante, se focaliza principalmente en la mencionada aplicación, puesto que el sistema central de IJOP está rodeado de oscuridad. No obstante, se ha podido averiguar que una de las fuentes de datos que emplea son las denominadas cámaras CCTV, algunas de las cuales tienen sistemas de reconocimiento facial con sensores de infrarrojos para las horas nocturnas, así como los llamados *wifi sniffers* (olfateadores de wifi), que captan los códigos y direcciones de identificación únicos de los ordenadores, *smartphones* y otros instrumentos. El sistema, además, recibe información desde innumerables sistemas sensoriales (denominados *checkpoints*) de la región, que reciben en tiempo real alarmas predictivas del IJOP para poder identificar candidatos que requieren control. Y no solo eso, el sistema, además, extrae información de ubicación de tales *checkpoints* para trazar el movimiento o trayectoria de las personas y asimismo se nutre de la información recopilada

por las empresas de transporte y de entrega de paquetes (nombre y dirección de la empresa remitente, quién recibe el paquete, si este fue recibido por la misma persona que constaba en el paquete o por un tercero, tipo de paquete, fecha y hora de firma de recepción del paquete, imágenes de rayos X del paquete y fotos del paquete de antes y después de su apertura, entre otros).

Aparentemente, el algoritmo empleado por el sistema IJOP para detectar ciudadanos que requieren control es sencillo: si A, entonces B (por ejemplo, si la persona que conduce un vehículo no es la misma que consta como titular, deberá investigarse).

De acuerdo con lo expuesto, el uso de tales herramientas en el modo en que se realiza en Sinkiang, obviamente conlleva un ataque frontal contra multitud de derechos fundamentales de sus ciudadanos, tales como derecho a la presunción de inocencia, el derecho a la libertad deambulatoria, el derecho a la intimidad, el derecho a la libertad religiosa, el derecho a la libertad de reunión, etc. Además, según establece un estudio publicado por la ya mencionada ONG Human Rights Wath en septiembre del 2018[788] las personas que son detenidas en Sinkiang no tienen derecho a la asistencia letrada, lo cual vulnera absolutamente todos los principios universales legales y éticos básicos, y en especial el artículo 12 de la Declaración Universal de Derechos Humanos ya anteriormente transcrito.

Así, en virtud de lo anteriormente puesto de manifiesto, puede advertir que, en caso de que los datos personales de los ciudadanos sean captados de forma masiva y lleguen a manos de personas u organizaciones con fines ética y legalmente cuestionables, como ocurre en el caso chino, el riesgo para los derechos de los ciudadanos es más que patente.

Y es que según las leyes internacionales de derechos humanos la vigilancia masiva no resulta legítima y, en España, en concreto, las investigaciones prospectivas se hallan absolutamente prohibidas, permitiéndose únicamente el seguimiento o la vigilancia de todos los movimientos de una persona en caso de que existan sospechas fundadas de que va a cometer un acto delictivo, previa autorización judicial en los términos previstos en la LECrim, con pleno respeto a lo dispuesto en la LO 7/21, de 26 de mayo, de protección de datos personales tratados

[788] Véase Human Rights Watch, 2018.

para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales.

No obstante, tal y como denuncia Amnistía Internacional, en algunos países se da carta blanca a tal tipo de vigilancia. Así, según se pone de manifiesto por tal organización, en Francia, por ejemplo, se permite interceptar comunicaciones de forma masiva y retener información durante largos periodos de tiempo, habiéndose prescindido de la autorizacion judicial previa; y, en Polonia, se han concedido a la policía y otras agencias gubernamentales poderes de vigilancia incompatibles con el respeto a la privacidad.[789]

En relación con ello, particularmente interesante es hacer mención a lo que ocurrió en Estados Unidos en 2013, cuando Edward Snowden hizo públicos unos documentos que revelaban cómo las agencias de seguridad de tal país llevaban a cabo vigilancias masivas para captar, almacenar y analizar en secreto millones de comunicaciones privadas de personas en todo el mundo, tal y como denuncia Amnistía Internacional.

Y es que fue el 5 de junio de 2013 cuando el exanalista de la CIA y de la National Security Agency (NSA) de EEUU Edward Snowden decidió denunciar públicamente, a través de los periódicos The Guardian y The Washington Post, la existencia de programas de vigilancia masiva de las comunicaciones de ciudadanos de todo el mundo. Así, el mencionado exespía manifestó que, sin control judicial alguno y bajo el paraguas de la seguridad, la NSA estadounidense y el gobierno de Reino Unido habían interceptado de forma masiva comunicaciones personales y comerciales de millones de ciudadanos (al parecer, incluso se espió el teléfono móvil de la canciller alemana Angela Merkel), tuvieran o no vínculo con la delincuencia. Además, hizo alusión a la existencia de una red de espionaje denominada Los Cinco Ojos[790] formada por los servicios de inteligencia de Estados Unidos, Canadá, Reino Unido, Australia y Nueva Zelanda que llevaban a cabo espionajes masivos y, aunque cumplían con las normas legales cuando intervenían las comunicaciones de sus propios ciudadanos, no actuaban igual cuando se trataba de ciudadanos extranjeros.[791]

[789] Véase Amnistía Internacional, s.f.
[790] En inglés, "*The Five Eyes*".
[791] Véase Campbell, 2013.

Y, asimismo importante es hacer alusión a la denuncia presentada ante el Tribunal Europeo de Derechos Humanos frente al gobierno ruso por la activista Alyona Popova y el político Vladimir Milov por el uso de tecnología de reconocimiento facial para llevar a cabo prácticas de vigilancia masiva durante una manifestación autorizada en Moscú que tuvo lugar el 29 de septiembre de 2019 por la exclusión de candidatos independientes de las elecciones legislativas de tal ciudad, siendo el primer caso de impugnación judicial del uso de tal tecnología con dichos fines.[792]

Igual de relevante es la denuncia que hizo Amnistía Internacional en relación a las noticias que surgieron en sendos medios de comunicación que aseguraban que las autoridades rusas habían detenido a varios asistentes a la manifestación pacífica que tuvo lugar el 21 de abril de 2021 en apoyo al activista Aleksei Navalny, tras su identificación con sistemas de reconocimiento facial. Respecto de ello, Natalia Zviagina, directora de la Oficina de Amnistía Internacional en Moscú expuso: "*Antes, el principal riesgo que corrían las personas que protestaban era que la policía las golpeara y las detuviera arbitrariamente en una concentración. Pero ahora evitar esta suerte no significa estar a salvo: el Estado represivo sabe quién eres y puede ir a buscarte en cualquier momento. El riesgo de tratar de ejercer tu derecho a la libertad de reunión pacífica nunca ha sido tan alto en Rusia*".[793]

Visto lo expuesto, y habida cuenta de la gran cantidad de información cruzada que los sistemas de vigilancia masiva pueden aportar (infinita más que la que podría detectar, captar y relacionar el ojo humano), está claro que ello es un fenómeno tentador, especialmente en materia de seguridad.

Y es que, la combinación y el cruce de aquellos datos obtenidos de forma tradicional y aquellos captados mediante herramientas basadas en IA puede arrojar enormes beneficios a la investigación de las causas. Así, por ejemplo, las huellas dactilares obtenidas en el lugar de los hechos pueden llevar a identificar a un sospechoso, que a su vez puede ser vinculado de forma automática con información relativa a sus últimos movimientos, al uso de sus tarjetas bancarias, a sus telecomunicaciones, viajes o actividades en las redes sociales, etc.

[792] Zoblina, 2020.
[793] Amnistía Internacional, 2021.

En relación con ello, procede hacer especial mención al proyecto VALCRI, financiado por la UE, que mediante el trabajo autómático y la colaboración con un equipo humano analiza los datos de una amplia gama de fuentes de formato mixto y presenta posibles explicaciones de los crímenes, lo cual allana el camino a la policía con argumentos rigurosos. Según se afirma en la página web del mencionado proyecto, VALCRI trabaja con inteligencia objetiva, velocidad y precisión para tratar de evitar el error y el sesgo humano y permite localizar rápidamente información interesante o sospechosa entre vastos conjuntos de datos. Y es que el programa analiza en tiempo real los datos entrantes junto con la información ya existente para ayudar a las autoridades a entender la relevancia y el interés que hay tras ello, pudiendo incluso recrear eventos delictivos utilizando reconstrucciones espacio-temporales. Sin duda, pues, es una herramienta interesante y con un potencial enorme cuyo éxito, no obstante, está supeditado a la buena calidad de la información empleada y a la legalidad de su tratamiento.[794]

En virtud de todo lo anteriormente expuesto, y ante un mundo cada vez más sujeto a vigilancias masivas (oficiales y no oficiales, legales e ilegales), están surgiendo aplicaciones de IA que tienen como objetivo salvaguardar la privacidad de los ciudadanos. Así, por ejemplo, resulta interesante hacer referencia a las denominadas "técnicas de desidentificación" de imágenes faciales, orientadas a proteger la privacidad de los ciudadanos y a eliminar los elementos o rasgos de identificación personal de las imágenes, dificultando así el uso fraudulento de los sistemas de reconocimiento facial, habida cuenta de que cuanto más fuerte es la anonimización aplicada a la imagen, mayor resulta la pérdida de utilidad y la disminución del valor de la misma.

Actualmente, los métodos de desidentificación de imágenes faciales incluyen, entre otras, técnicas de pixelación y desenfoque, disminución de la calidad, etc, lo que, evidentemente, reduce muy notablemente las posibilidades de éxito de los sistemas de reconocimiento facial. Y ello lo que hace es evitar, justamente, que las imágenes de los rostros de los ciudadnos sean víctimas de posibles intromisiones no autorizadas por parte de sistemas de reconocimiento facial automático, por ejemplo. En concreto, D-ID™ es una tecnología que emplea métodos de modificación de imágenes específicamente diseñados conservar la naturalidad visual de las mismas, de modo que la representación de la cara modificada es casi idéntica a la original

[794] Véase VALCRI, s.f..

pero resulta resistente a los sistemas de reconocimiento facial, lo que la protege eficazmente del mal uso de los mismos.[795]

En relación con ello, entre otras, la *startup* con sede en Berlin llamada Brighter AI ha lanzando un producto denominado "*ProtectPhoto*" que anonimiza los rostros en las imágenes mediante la extracción de las caras originales y el reemplazo por otros nuevos que no se pueden rastrear. Y la misma compañía, además, comercializa tecnología *deepfake* que permite difuminar matrículas y desenfocar rostros para que las empresas o gobiernos que emplean cámaras de videovigilancia cumplan con las leyes de privacidad europeas, especialmente con el RGPD.[796]

No obstante, en la actualidad, las prácticas de vigilancia masiva son cada vez más sencillas, no solo por la proliferación de la instalación de cámaras de videovigilancia (en ocasiones a tiempo real) en la vía pública, el uso de los *smartphones* y de Internet, entre otros, sino también por la creciente utilización de los dispositivos conectados en los hogares, los vehículos y las personas (por ejemplo, elementos de domótica, relojes o pulseras inteligentes) que, gracias a la combinación de la tecnología 5G y la denominada IA de las cosas (en inglés, *AI of Things*), está permitiendo la creación de una cotidianeidad conectada.

Hoy en día no existe en la UE una regulación como tal, específica, sobre las prácticas de vigilancia masiva, si bien deben ser de aplicación las distintas normativas relativas a cada una de las fuentes de información disponibles (a saber, normativa sobre videovigilancia, normativa sobre interceptación de las comunicaciones, normativa sobre protección de datos personales, etc) teniendo en cuenta siempre, de modo principal, la prohibición que rige sobre las investigaciones prospectivas y el respeto a los derechos fundamentales de los ciudadanos. Así, en caso de que a un juez de instrucción le llegara una *notitia criminis* que tuviera origen en una práctica de vigilancia masiva, lo más relevante sería, sin duda, determinar su legitimidad.

En España, no obstante, y a modo de ejemplo, ya existen proyectos que combinan distintas técnicas de IA para controlar de forma masiva, por ejemplo, eventos multitudinarios. En

[795] Véase D-ID, 2019.
[796] Véase Koetsier, 2020.

relación con ello es interesante hacer especial mención a lo previsto en la Resolución de 16 de julio de 2020, de la Subsecretaría de Estado de Interior, por la que se publica el Convenio entre el Centro para el Desarrollo Tecnológico Industrial, E.P.E., y el Ministerio del Interior, relativo a la contratación precomercial de servicios de I+D en materia de seguridad en el medio rural. En concreto, en el Anexo I, se contempla el "*Escenario 2: Control de eventos multitudinarios*", y en relación a ello se dispone:

"*Las soluciones innovadoras propuestas por la industria para este escenario, además de los aspectos comunes, deben tener en cuenta los siguientes aspectos específicos:*

Área 1: Identificación de personas con asuntos pendientes con la justicia en los puntos de control de acceso al evento.

-Se prevé que la solución innovadora pueda basarse, de manera general, en un sistema tecnológico innovador e inteligente formado por los siguientes componentes:

• *Sistema de reconocimiento de matrículas. A ser instalado de manera previa a la entrada del aparcamiento del evento (por ejemplo, podría ser instalado en postes fijos o en vehículos aéreos no tripulados). Su propósito es proporcionar a los agentes que controlan los accesos al aparcamiento alertas para detener a vehículos con asuntos pendientes con la justicia.*

• *Sistema de detección de teléfonos móviles (basado en IMSI-catcher). A ser instalado en el propio punto de control de acceso en la entrada al evento. Su propósito es proporcionar a los agentes que controlan los accesos al evento alertas para detener a personas con asuntos pendientes con la justicia.*

• *Sistema de reconocimiento de personas (por ejemplo, sistema de reconocimiento facial). A ser instalado en el propio punto de control de acceso en la entrada al evento. Su propósito es proporcionar a los agentes que controlan los accesos al evento alertas para detener a personas con asuntos pendientes con la justicia.*

• *Sistema/s de procesamiento de datos y generación de alertas.*

• *Sistema/s de visualización de información y alertas.*"

3.3. HERRAMIENTAS DE TRAMITACIÓN -breve reseña-

Es evidente que el éxito de la instrucción de las causas viene dado, en gran parte, por su contenido, pero lo cierto es que la eficiencia en la tramitación resulta fundamental para poder conseguir dar un buen servicio, de calidad, al ciudadano, que es a lo que principalmente se debe aspirar desde la Administración de Justicia.

Y es que por muy brillante que resulte una investigación policial e incluso judicial, si el caso no tiene el impulso debido lo más probable es que la fase de instrucción se eternice y, lo que en un principio podía resultar relativamente sencillo, se complique exponencialmente por razones varias: las víctimas ya no recuerdan el rostro del autor de los hechos, un testigo se va a vivir al extranjero, el investigado se cambia de domicilio sin comunicarlo al Juzgado y debe ser llamado por requisitorias, el juez de instrucción y/o el fiscal cambian en más de una ocasión por haber movimiento de destinos, se interponen más recursos, etc.

No obstante, lo cierto es que, desde que entró en vigor el artículo 324 LECrim, el control sobre los plazos de la instrucción de las causas se ha incrementado de forma notoria. Sin embargo, y por desgracia, sin duda alguna la eficiencia en la Administración de Justicia sigue siendo una asignatura pendiente. Y es que hay mucha rotación de personal, la formación (especialmente práctica) que se otorga a los operadores jurídicos es insuficiente, hay un nivel excesivo de interinidad (y en no pocas ocasiones acceden a puestos de tramitación y gestión, por ejemplo, personas que no han trabajado nunca antes en el ámbito de la justicia), el volumen de asuntos que entra en los juzgados excede con creces los límites de capacidad de trabajo del personal, los sistemas informáticos fallan con frecuencia, etc.

Tal nefasta situación (que especialmente se da en los Juzgados mixtos de la periferia pero que está extendida por todos los órganos judiciales de España), no obstante, sin duda alguna podría verse mitigada con el uso de la IA para dar soporte a la gestión y la tramitación de las causas, puesto que esta no entiende de sobrecarga de trabajo, de desactualizaciones, de traspapeleo, de días de permiso, de vacaciones, etc. No obstante, como es lógico, entiendo que ese reparto

de asuntos siempre debería poder ser revisado por un humano cualificado, pero ya con un filtro previo que, con toda seguridad, le facilitaría mucho el trabajo.

Y es que imaginemos lo útil que resultaría emplear un sistema informático que, mediante el uso de la IA, tras el registro de entrada, pudiera clasificar las denuncias de particulares, las querellas, los partes médicos, los atestados, los informes de centros penitenciarios, etc y filtrar la competencia por razón del lugar y la fecha de comisión de los hechos, del carácter ampliatorio de las diligencias, de la existencia de un turno especial de reparto o de la existencia de antecedentes, siendo que los algoritmos utilizados estarían entrenados para ejecutar tal tarea con base en las normas de la LECrim y las específicas reglas de reparto del correspondiente partido judicial.

Además, una vez determinada la competencia, el sistema de IA asimismo podría indicar a los funcionarios cuáles serían los siguientes pasos a seguir, a saber: en caso de que una querella no estuviera presentada por Abogado y Procurador, requerir por X días para subsanar con apercibimiento de inadmisión a trámite, por ejemplo; y al juez, a saber: en caso de tratarse de una denuncia por estafa de más de cuatrocientos euros, incoar Diligencias Previas y acordar la declaración del denunciante como perjudicado y hacerle el ofrecimiento de acciones, acordar oficiar a tal banco o a tal otro para que remitiera la información necesaria, por ejemplo, y acordar la declaración del denunciado como investigado, previa recabación de sus antecedentes penales; o, en caso de tratarse de una estafa de menos de cuatrocientos euros, incoar juicio por Delito Leve y señalar vista con citación de los denunciantes, en su caso, los testigos, y los denunciados.

Y es que efectivamente los algoritmos del sistema podrían estar entrenados para sugerir al juez de instrucción las diligencias a practicar en función del delito de que se tratara, y luego ya, claro está, sería este quien decidiría cuáles de ellas llevar a cabo y cuáles no. Además, el sistema podría estar diseñado para que, conforme se fueran practicando las mencionadas diligencias, estas fueran siendo marcadas como “practicadas” y, a medida que se acercara el fin del plazo de instrucción previsto en el artículo 324 LECrim, generara una alerta para decidir sobre la necesidad de prórroga y sobre la necesidad de la práctica de las diligencias todavía no practicadas.

Tal sistema podría ayudar, asimismo, a la admisión o inadmisión a trámite de los recursos que se presentaran, verificando el tipo de recurso legalmente previsto para cada clase de resolución y el plazo de interposición. Y es que, por ejemplo, en caso de que contra un Auto de continuación de Procedimiento Abreviado se interpusiera en forma por la defensa, dentro del plazo legal, un recurso de reforma y subsidiario de apelación, este podría ser admitido a trámite directamente, con traslado al Ministerio Fiscal y a las partes personadas para que formularan alegaciones en el plazo legalmente previsto, y una vez recibidas, podría darse cuenta a SSª para resolver. O en caso de que dicho recurso se presentara fuera de plazo, podría darse aviso automático al juez para que procediera a su inadmisión a trámite.

Interesante sería, además, que tanto a los funcionarios como al Ministerio Fiscal y al juez de instrucción les fueran saltando alertas diarias por las tareas pendientes cuando fueran aproximándose las fechas límite para resolver, para que así estos pudieran dar prioridad a lo más urgente e intentaran ceñirse, en la medida de lo posible, a los plazos legalmente previstos.

Y es que ello, sin duda, ayudaría a optimizar los recursos de la Administración de Justicia y permitiría tanto a los funcionarios como a los fiscales y a los jueces de instrucción dedicar más tiempo a llevar a cabo tareas que requieren un mayor componente intelectual humano, lo cual sería altamente beneficioso para el ciudadano.

Además, reduciría los efectos de las negligencias, del mal funcionamiento de la Administración de Justicia por falta de recursos y, en concreto, entre otros, disminuiría la apreciación de la hoy generalizada circunstancia atenuante de dilaciones indebidas, rebajaría el número de procedimientos que acaban archivados por prescripción, y, en resumen, elevaría los ratios de éxito de la tarea instructora y aminoraría el sufrimiento y la ansiedad que provoca a muchos ciudadanos la larga espera que media entre el inicio de la instrucción de las causas y su enjuiciamiento. No obstante, sin duda, y para que tales sistemas tuvieran un impacto real, deberían ser asimismo empleados por los órganos de enjuiciamiento, puesto que sino estos, en caso de mantener el sistema de tramitación tradicional, no podrían absorber el elevadísimo número de procedimientos que les irían llegando de parte de los órganos instructores.

No obstante, procede poner de manifiesto que según un informe elaborado por la Comisión Europea para la Eficiencia de la Justicia[797], el uso de IA en el poder judicial parece ser menos popular en Europa que en Estados Unidos. Y es que en tal país ya se emplean numerosos *software* de apoyo a la tramitación y la gestión de los expedientes judiciales, entre otros, por ejemplo, JWorks, Court View y ShowCase, comercializados por la empresa Equivant y, sin embargo, en el ámbito de la UE, el uso de este tipo de herramientas no está todavía extendido.

En relación con ello, y a nivel internacional, no obstante, es especialmente relevante hacer mención al caso de Argentina, que cuenta en mi opinión, con el sistema estrella de IA de gestión y tramitación procesal: Prometea.

Y es que Prometea, un sistema experto con multiplicidad de funciones que, entre otras técnicas, emplea *Machine Learning*, nació en 2017, sin precedentes, en el ámbito del Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires y fue ideado con el objetivo de agilizar los procesos judiciales y optimizar el servicio de la Administración de Justicia, si bien hoy en día se halla en pleno rendimiento en otras instituciones, tanto en la Ciudad de Buenos Aires como en otras provincias, y está en marcha su implementación en la Corte Interamericana de Derechos Humanos y en la Corte Constitucional de Colombia.

Según se informa en la página web del Ministerio Público Fiscal de Argentina, Prometea se caracteriza por tres grandes rasgos:

"*a) posee una interfaz intuitiva y amigable que permite "hablarle" al sistema o chatear a partir de un reconocedor de lenguaje natural;*

b) opera como sistema experto con multiplicidad de funciones, que permite automatizar datos y documentos y realizar asistencia inteligente;

c) utiliza técnicas de machine learning supervisado y de clustering, a partir de etiquetado manual y de máquina con dataset de entrenamiento, a efectos de realizar predicciones y/o detecciones en grandes volúmenes de documentación."

[797] Véase Consejo de Europa, 2018.

Y es que, según se desprende de la mencionada página web, Prometea tiene capacidad para, en pocos segundos, realizar informes, segmentar y filtrar documentación en función de su contenido, buscar información en archivos, elaborar indicadores con gráficos comparativos, proporcionar respuestas de manera automática a partir de un determinado *input*, etc.

Y, en relación con ello, según estudios realizados respecto del mencionado sistema, ha quedado demostrado que este es capaz de reducir los tiempos de realización de las tareas judiciales hasta en un 90%, además de minimizar (e incluso en algunos casos, directamente eliminar) cualquier margen de error.

Respecto de la eficiencia de la herramienta, en la ya citada página web se pone de manifiesto que, entre otros beneficios:

-predice las soluciones de los casos judiciales en menos de veinte segundos, con una tasa de acierto igual o superior del 90%;
-contrasta información entre documentos y bases de datos sin intervención humana y logra reducir tiempos de trabajo de trece horas a cinco minutos;
-elabora una base de datos de forma automática a partir de la extracción de datos de la documentación de que dispone;
-filtra, sistematiza y clasifica informes en pocos segundos, en función de su contenido, en investigaciones de abuso sexual infantil, lo cual conllevaría más de ocho horas de trabajo para un ser humano;
-elabora ciertas resoluciones de manera automática, lo que reduce el tiempo de confección, en el caso de los Decretos de determinacin de hechos, por ejemplo, de veinticinco minutos a tan solo dos; y
-reduce los tiempos de redacción de sentencias hasta en un 80%, en ciertas oficinas judiciales.

Como consecuencia de las mencionadas bonanzas, la herramienta Prometea ha sido presentada en diversas instituciones tanto nacionales como internacionales, habiendo sido reconocida como un caso de éxito mundial en el ámbito de la Justicia por la publicación N°

31 de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) sobre Gobernanza Pública.[798]

Y todo ello ha sido contrastado con Daniela Dupuy, Fiscal Jefa de la Unidad Fiscal Especializada en Delitos y Contravenciones Informáticas de la Ciudad de Buenos Aires (Argentina), quien de primera mano me expuso que Prometea ha supuesto una revolución para los fiscales y los jueces argentinos, habida cuenta de que es muy flexible y se adapta a cualquier proceso en función de sus necesidades; su uso puede ser combinado con los sistemas de gestión, lo que multiplica su capacidad de auxilio; es perfectamente transparente, puesto que la trazabilidad de sus tareas está garantizada; etc, y además los desarrolladores (ingenieros), en relación constante con los jueces y fiscales, trabajan sin cesar para mejorar el sistema y lograr nuevas utilidades con el fin de optimizar recursos y proporcionar el mejor servicio público de justicia.

En España, no obstante, el uso de tales sistemas de IA de forma generalizada todavía es una utopía, máxime siendo que, además, la competencia en medios materiales en materia de Justicia está cedida a las Comunidades Autónomas, en el marco de lo previsto en los artículos 148 y 149 de la Constitución Española, por lo que en cada una de ellas se emplea un sistema distinto de tramitación, más o menos moderno.

No obstante, en mi opinión, las CCAA deberían tratar de aunar esfuerzos en un único sentido e, incluso, intentar concertar acuerdos de colaboración con sus homónimos argentinos, que tengo constancia de que estarían dispuestos a compartir su tecnología para uso judicial, con adaptación de sus aplicaciones, eso sí, a los estándares de protección de datos personales que rigen en la UE. Y es que, desde luego, la combinación de un sistema del nivel de Prometea con los actuales modelos digitales de tramitación y gestión procesal, podría incrementar enormemente la eficiencia de la Administración, y siendo que la justicia es un servicio público que debe aspirar a la calidad y al trato igual a todos los ciudadanos, con independencia del territorio en el que estos residan, un paso al frente de tal calado, en materia tecnológica, podría resultar la revolución sin precedentes que todos llevamos décadas esperando.

[798] Véase Ubaldi, 2019.

4.- CONCLUSIONES

El presente trabajo, tal y como ya se anunció en la Introdución y tal y como se ha ido reiterando en su transcurso, tenía como principal finalidad la de investigar qué utilidades puede tener la IA en el proceso de instrucción español con el fin de mejorar su eficiencia y, por ende, la calidad del servicio que la Administración de Justicia presta al ciudadano, sin obviar sus potenciales riesgos.

Para ello, ha resultado imprescindible no solo realizar un estudio inicial sobre el concepto de IA y sus distintas aplicaciones, sino también efectuar un profundo análisis sobre la relación existente entre tal tecnología y el Derecho, habida cuenta de los enormes riesgos que esta entraña y que desde el ámbito jurídico procede, en todo caso, mitigar. Y, ya sentadas las antedichas bases, se ha procedido a llevar a cabo un exhaustivo examen de todas aquellas herramientas de IA que podrían resultar de utilidad en la instrucción penal de las causas teniendo en consideración, evidentemente, sus potenciales riesgos, haciendo referencias en cada caso a la regulación vigente y/o venidera.

En cualquier caso, entiendo necesario advertir que la presente tesis doctoral no constituye, ni mucho menos, un puerto de llegada, sino de partida, habida cuenta de que las cuestiones que en ella se abordan son muy novedosas y, sin duda, resultarán objeto de múltiples ulteriores y, en ocasiones, discrepantes estudios (especialmente cuando, por fin, se publique el Reglamento europeo sobre IA). No obstante, ello me parece un escenario deseable y "sano", puesto que será el mejor indicador de que la investigación criminal evoluciona, y ese es esencialmente el espíritu que, como servidora pública, me gustaría transmitir: mirar hacia adelante con voluntad de mejora. Siempre.

Expuesto cuanto antecede, procede poner de manifiesto que la realización de este trabajo me ha llevado a extraer las conclusiones que, a modo de propuesta personal (sabedora de que pueden existir otros puntos de vista), paso a exponer.

PRIMERA.- La IA puede reportar grandes beneficios y utilidades al proceso penal de instrucción español, pero también enormes riesgos.

Una de las más aplaudidas y, bajo mi punto de vista, acertadas frases célebres atribuidas al tecnólogo por excelencia de nuestra era, Bill Gates, por el nivel de optimismo pero a la vez de cautela que lleva implícitos, es la que sigue: "*La primera regla de cualquier tecnología utilizada en una empresa es que la automatización aplicada a una operación eficiente magnificará la eficiencia. La segunda es que la automatización aplicada a una operación ineficiente magnificará la ineficiencia*".

Y es que considero que tal aseveración, cargada de contenido y fuerza, aplicable, sin duda, al uso de la IA tanto en el sector público como en el sector privado, podría ser el perfecto resumen de lo que he tratado de exponer a lo largo de la presente tesis doctoral. Me explico.

Tal y como puede deducirse del contenido de este trabajo, la IA, desde luego, es una herramienta tecnológica sin precedentes que cuenta con capacidad más que suficiente para magnificar la eficiencia de las tareas humanas. No obstante, ello tiene un importante matiz. Y es que, como se ha ido explicando, la mencionada tecnología funciona a través de potentes sistemas de *software* que llevan a cabo el análisis de ingentes cantidades datos para arrojar unos resultados determinados. Así, su éxito o su fracaso depende, por un lado, de la potencia y de las características técnicas de tales sistemas y, por otro lado, fundamentalmente, de la calidad de los datos analizados.

De acuerdo con ello, por ejemplo, si en una organización, ya con anterioridad a la introducción de la IA, se tomaban decisiones humanas, no automatizadas, con base en datos de mala calidad (a saber, obtenidos de forma ilícita, no representativos, erróneos, etc) y, por ende, incorrectas e ineficientes (habida cuenta del incompetente y poco útil empleo de recursos), la incorporación de una tecnología tan poderosa como la IA y la automatización de las tareas sin introducir cambios en la calidad de tales datos no hará más que multiplicar y magnificar la ya histórica ineficiencia de la organización y de sus operaciones, lo cual podría acarrear resultados altamente perjudiciales para los ciudadanos, habida cuenta de que se perpetuarían patrones que, sin duda, deberían tender a desaparecer.

No obstante, esta aparentemente sencilla reflexión alberga un amplio debate técnico y jurídico que, especialmente en relación con el uso de la IA en el ámbito de la investigación penal (y,

en concreto, en el proceso de instrucción español), es el que ha dado base a la presente tesis doctoral.

Ya al comienzo de este trabajo procuré advertir de la gran dificultad que entraña establecer una definición universal y permanente de lo que es la IA sin riesgo de caer en la obsolescencia, por la veloz y constante evolución a la que tal tecnología se haya sometida. No obstante, cierto es que a lo largo del mismo se han ido sentando las bases del antedicho concepto y se han ido perfilando los elementos comunes a todas aquellas técnicas que conforman la mencionada noción.

Y es que, sin duda, entiendo que una de las cuestiones que han quedado claramente patentes tras examinar las distintas aplicaciones que tiene o puede tener la IA en el ámbito de la investigación criminal, es que por mucho alboroto terminológico y técnico que subyazca, la realidad es que hay un elemento común en todas ellas que es el que sirve de fundamento e hilo conductor, principalmente, para los que somos legos en materia tecnológica, a saber: la detección y el reconocimiento de patrones entre ingentes cantidades de datos como base para la toma automática de decisiones.

Así, la razón de ser de la IA no es (o no debería ser, a mi entender) otra que la de la resolución de problemas mediante el uso de la máxima información posible (a lo que habría que añadir: "de calidad"), para llevar a cabo la adopción de las mejores y más rápidas decisiones en beneficio de los seres humanos.

No obstante, tal y como se ha ido poniendo de manifiesto a lo largo del presente trabajo, por desgracia, ello no siempre es así.

Y es que, las bonanzas que, en general, los sistemas de IA pueden aportar a nuestras sociedades están más que claras, a saber, y de modo principal: la liberación para los humanos de tareas automatizables y, por ende, el incremento de su tiempo y su capacidad para llevar a cabo actividades que requieren cualidades esencialmente humanas (valga la redundancia), a saber, la creatividad o la empatía; y el aumento (a niveles en ocasiones inimaginables) de la eficiencia -fundamentalmente en términos de precisión, tiempo y costes- en la toma de decisiones.

En concreto, en el ámbito de la investigación criminal, los posibles beneficios son, asimismo, patentes, a saber, entre otros: el incremento de la eficiencia en la prevención delictiva; el auxilio a las autoridades -policía judicial, fiscales y jueces- en la toma de decisiones basadas en la predicción de hechos futuros (especialmente en la adopción de medidas cautelares); el aumento de la eficiencia de los recursos materiales y personales (lo cual se traduce, esencialmente, en un incremento de la calidad y de la rapidez de la resolución de los casos y en una disminución de los costes); y el incremento sin precedentes de las tasas de éxito de la instrucción de las causas.

No obstante, tal y como ya se ha apuntado en varias ocasiones a lo largo de esta tesis doctoral, los potenciales riesgos que el uso de los sistemas de IA entrañan no pueden perderse de vista en ningún momento, habida cuenta de que estos tienen capacidad para superar con creces a los beneficios enumerados y ejercer un efecto multiplicador de las históricas debilidades presentes tanto en el proceder policial como en el judicial, tal y como advierte Bill Gates en la célebre frase que da comienzo a la presente Sección.

Así, con carácter general, considero que los humanos debemos estar atentos a la posible presencia de, entre otros, los siguientes riesgos: la vulneración masiva de derechos fundamentales y libertades individuales como consecuencia del uso indiscriminado de sistemas de IA; la pérdida del control humano sobre los sistemas de IA y, por ende, la potencial quiebra del principio de uso de estos en beneficio de la humanidad; el incremento de la brecha entre las mayorías y las minorías y los países ricos y aquellos en vías de desarrollo; la ausencia de transparencia y explicabilidad de los sistemas de IA y, por ende, la pérdida de derechos y de confianza de los ciudadanos; las dificultades técnicas y la existencia de brechas de seguridad que provoquen consecuencias irreparables para los seres humanos; los problemas relacionados con la responsabilidad por el eventual mal uso o mal funcionamiento de los sistemas de IA y la consiguiente desprotección de los ciudadanos afectados; o la pérdida de puestos de trabajo y el inevitable empobrecimiento de la sociedad. Y, en relación con esto último, resulta muy gráfico el ejemplo puesto ya en 2015 por el científico Stephen Hawking, que advirtió: "*Los robots podrían hacernos ricos y libres, pero es más probable que terminemos pobres y desempleados*".[799]

[799] Véase Bolton, 2015.

En concreto, en el ámbito de la investigación criminal, además, debemos estar atentos especialmente, a una posible transgresión de derechos fundamentales tan esenciales como la dignidad de la persona, la libertad, la igualdad y la no discriminación, el derecho a la defensa y el derecho al honor, a la intimidad y a la protección de datos personales, entre otros; un potencial empleo de los sistemas de IA para llevar a cabo la perpetración de delitos de forma más sofisticada e impune; una limitación del derecho de acceso a la justicia en función del grupo o país al que pertenezca cada individuo; la imposibilidad de conocer si se ha llevado a cabo un proceso legal y con todas las garantías, de conocer si se han producido vulneraciones de derechos, y de rebatir los argumentos que han llevado a las autoridades policiales, fiscales y/o judiciales a tomar determinadas decisiones, por falta de transparencia; el desamparo de los ciudadanos afectados por el mal uso o funcionamiento de los sistemas de IA; el aumento del desempleo entre los funcionarios de Justicia y las autoridades policiales, fiscales y judiciales; y, sobre todo, lo más preocupante: la deshumanización de la justicia.

Y es que tales riesgos que, ya de por sí, en general, entrañan graves peligros para los seres humanos, en el campo de la investigación penal se multiplican y resultan absolutamente intolerables y contrarios a todos los principios constitucionales y legales vigentes en nuestro país, tal y como ya he ido poniendo de manifiesto a lo largo de la presente tesis doctoral.

SEGUNDA.- La única forma de que la IA pueda ser empleada por parte de las autoridades en el ámbito de la instrucción penal sin llevar a cabo vulneraciones masivas de derechos es su regulación.

Así, tras analizar de forma profunda los peligros expuestos, he llegado a la conclusión de que la única solución posible para emplear la IA en el ámbito de la investigación penal de forma garantista y segura es la de tratar de minimizar tales riesgos mediante la regulación. Una regulación que, por un lado, debe definir y delimitar muy bien en qué casos pueden ser empleados los sistemas de IA de investigación criminal, y en qué términos; por otro lado, debe establecer mecanismos para garantizar la calidad de los mismos tanto en el momento de su comercialización o puesta en circulación como con posterioridad; y, finalmente, debe contener elementos de control del cumplimiento de los requisitos legalmente exigidos en cada momento.

Ya en el siglo XIX el jurista alemán Federico Carlos de Savigny, convencido romanista, advirtió de los inconvenientes y problemas que podía provocar la codificación, habida cuenta de que la realidad social es altamente cambiante y modificar las leyes no es una tarea ni ágil ni ligera, por lo que este entendía que la regulación podía implicar que el Derecho fuera siempre un paso (o más) por detrás de la realidad. Tal argumento, sin duda, en mi opinión, resulta en buena parte acertado, y más en un ámbito como el de la IA, que varía y evoluciona a mucha velocidad. No obstante, considero que la visión del también jurista alemán Anton Friedrich Justus Thibaut, que entendía necesario poner orden en la legislación, completarla y sentar unas bases jurídicas claras para acabar con el, bajo mi punto de vista peligroso "particularismo jurídico", y dar paso a la seguridad jurídica, es asimismo idónea.

Y es que ya hemos visto a lo largo del presente trabajo que el guirigay legislativo, incompleto e insuficiente al que nos enfrentamos en la actualidad al hablar de IA, especialmente en el ámbito de la UE (siendo que no existe una regulación específica y clara), no lleva más que a tratar de adivinar qué instrumento jurídico resulta de aplicación en cada caso, muchas veces sin éxito, lo que abre la puerta a la existencia de vacíos legales y situaciones confusas que no hacen más que causar desconfianza y desprotección a los ciudadanos.

Así, bajo mi punto de vista, una condición previa indispensable para llevar a cabo el uso de herramientas de IA en cualquier caso y, especialmente, en el ámbito de la investigación de delitos, es la de legislar, habida cuenta de que es la única forma de generar seguridad jurídica y confianza en los ciudadanos y, sobre todo, de evitar vulneraciones masivas de derechos fundamentales. Y entiendo que lo más idóneo, en tal sentido, y en aras de actuar en equilibrio entre lo defendido tanto por Savigny como por Thibaut, sería establecer una regulación "de mínimos", que fuera clara y específica pero que dejara la puerta abierta a examinar cada caso concreto de forma flexible, justamente por la constante y rápida evolución que está experimentando la IA.

En la actualidad, no obstante, en la UE y en España, por desgracia, ello no ha tenido lugar, lo cual ha dejado la puerta abierta a numerosas violaciones de derechos y libertades que, de forma constante, han ido siendo denunciadas por las organizaciones *pro* derechos humanos que "sacan los colores" a las autoridades en la mayoría de ocasiones con toda la razón.

Ante tal problemática, no obstante, y ante las barbaridades que se han estado cometiendo durante los últimos años (y continúan cometiéndose) mediante el uso de sistemas de IA sin contar con una regulación clara y propia, se ha ido tomando conciencia por parte de las autoridades, por fin, de la necesidad de establecer un marco jurídico claro aplicable al diseño, al desarrollo, a la comercialización, a la distribución y al uso de los sistemas de IA.

TERCERA.- La necesaria elaboración de una Declaración Universal de Principios sobre IA.

En relación con ello, y con independencia de que cada país (o conjunto de países, como en el caso de la UE), en el ejercicio de su propia soberanía, proceda a elaborar una legislación específica sobre tales extremos (que deberá plasmar los principios y valores que el poder legislativo considere oportunos en cada momento), tal y como ya he puesto de manifiesto a lo largo de este trabajo, entiendo que resultaría muy necesario y beneficioso para la humanidad que se establecieran unos principios básicos de IA, una especie de "*Bill of Rights*", de forma similar a lo que ya se hizo el 10 de diciembre de 1948 con la Declaración Universal de Derechos Humanos, que garantizara unos mínimos éticos y morales que los sistemas que emplean tal tecnología debieran cumplir a nivel internacional y se erigieran como un verdadero escudo de protección frente a esta nueva, envolvente y peligrosa tecnología.

Para ello, no obstante, se necesitaría el marco de una organización internacional como la ONU o cualquier otra institución pública, preexistente o creada *ad hoc*, que consiguiera reunir al mayor número de países firmantes posible y, deseablemente, lograra hacer que tal declaración de principios ostentara carácter vinculante, estableciendo las correspondientes sanciones para casos de incumplimiento y fijando un sistema de prevención y disuasión infractora altamente potente, ya que hay que tener en cuenta que en caso de vulneración de las bases acordadas, recuperar el control y depurar responsabilidades de forma efectiva resultaría prácticamente imposible.

Y ello, sin duda, en mi opinión, es una mera cuestión de voluntad y de compromiso. Voluntad de ceder un poco de soberanía nacional propia a corto plazo a cambio de conservarla como especie en el futuro; voluntad de limitar la productividad y la riqueza nacional a cambio de mantener la libertad como conjunto de seres humanos; y compromiso de no romper las reglas

en beneficio de todos, con el fin de garantizar la supervivencia. Y es que, bajo mi punto de vista, si no se fijan a nivel mundial unos principios básicos claros y vinculantes que guíen a todos en el uso de la IA, el riesgo de que tal tecnología acabe fuera del control humano y nos desnaturalice como especie es real. Y estamos advertidos.

En relación con ello, es evidente que no todos los países tienen los mismos intereses y la misma facilidad para llegar a acuerdos con otros (ya que algunos de ellos son incluso enemigos), pero en este caso yo opino que los mandatarios de todo el mundo deberían hacer un esfuerzo extra y mirar "más allá de sus narices" para adoptar compromisos de calado internacional que, sin duda, revertirían en su propio beneficio en el futuro, ya que la alternativa bajo mi punto de vista es extremadamente peligrosa. Y lo peor es que basta con que un país con cierto poder haga mal las cosas para que el resto (y él mismo) puedan verse perjudicados, por lo que el pacto debería, en mi opinión, aspirar a ser firmado por la totalidad de los países existentes, lo cual, además, implicaría una mínima garantía de igualdad que podría llevar a vencer la actual brecha que acecha hoy en día al mundo (especialmente al continente africano, que se halla a la cola) en materia de IA.

No obstante, por desgracia, y a pesar de la gran cantidad de iniciativas y proyectos que hay en marcha para regular la IA, lo expuesto se antoja todavía remoto, ya que no parece que, al menos por el momento, puedan tenerse expectativas reales sobre la creación de una Declaración Universal de Principios de IA que ostente carácter vinculante para la práctica totalidad de los países del mundo, debiéndonos conformar con meras declaraciones de principios con valor orientativo que, esperemos, sirvan al menos para inspirar a los poderes legislativos nacionales (o transnacionales, como es el caso de la UE) a la hora de establecer una legislación propia sobre IA.

De todas formas, está claro que estamos ante un fenómeno sin precedentes y, afortunadamente, la concienciación de la sociedad sobre la amenaza que la IA supone y la necesidad de establecer límites a la misma está creciendo, por lo que me resisto a descartar la opción expuesta.

Y en relación con ello, tal y como ya he manifestado a lo largo del presente trabajo, entiendo que los principios básicos que deberían contemplarse en la mencionada propuesta de Declaración Universal de Principios de IA serían los siguientes:

- Principio de respeto a la dignidad del ser humano, con garantía de supervisión y control, así como de prioridad del bienestar social y ambiental;
- Principio de respeto a la libertad y a la privacidad del ser humano, con garantía de gestión individual de los datos personales, transparencia y explicabilidad de los sistemas;
- Principio de equidad, igualdad, no discriminación del ser humano e inclusión;
- Principio de robustez, solidez técnica y seguridad; y
- Principio de responsabilidad.

CUARTA.- La imperante necesidad de elaborar leyes nacionales (o supranacionales) que garanticen un uso responsable y garantista de la IA en el ámbito de la investigación penal.

Sentado lo anterior, a pesar de no contar actualmente, por desgracia, con un marco jurídico básico común, entiendo que resulta urgente que los distintos países procedan a regular la IA en el ámbito de su propia soberanía o de la de las organizaciones de las que formen parte (como por ejemplo, la UE). Y es que considero que, hasta ahora, el desconocimiento y el temor que se tenían hacia la mencionada tecnología, unidos a la esperanza de que se fijaran a nivel internacional unos principios comunes de actuación, han sido la principal causa de la pasividad de los Estados respecto del establecimiento de una legislación nacional clara y específica de tal tecnología. No obstante, actualmente, siendo que el uso de la IA está a la orden del día y la ausencia de regulación está provocando en muchas ocasiones vulneraciones masivas de derechos y libertades de los ciudadanos, como ya se ha dicho, la necesidad de legislar es apremiante. Ya vamos tarde.

Especialmente crítica soy en tal sentido con la UE, que si bien, tal y como he expuesto en la presente tesis doctoral, en los últimos tiempos está dando pasos firmes hacia la regulación de la IA, lo cierto es que ha estado "dormida" durante una época clave y, como consecuencia de ello, se ha visto superada por países como EEUU, China o Israel en relación al uso de tal

tecnología, lo cual me parece especialmente grave, puesto que entiendo que la UE, el mercado único más grande del mundo, ostentaba la responsabilidad de elaborar una regulación conforme con sus principios y valores que garantizara un empleo de la IA respetuoso con los derechos fundamentales.

Y es que el poder de influencia de la UE es enorme y, sin duda, traspasa sus fronteras, habida cuenta de que los países productores de tal tecnología (que, por lo general, suelen ser aquellos menos respetuosos con los derechos humanos), en caso de que desearan comercializar sus productos en el ámbito de la Unión, deberían adaptarse a los estándares fijados por esta, lo cual ya les disuadiría de entrada, en gran medida, de elaborar sistemas que no comulgaran con los valores y principios de la UE, altamente garantistas para los derechos y las libertades de los ciudadanos.

No obstante, mi decepción va todavía más allá, ya que en los propios Estados Miembros de la UE se han venido empleando herramientas de IA que, como ya he advertido con anterioridad, han sido denunciadas en reiteradas ocasiones por organizaciones *pro* derechos humanos por sus potenciales vulneraciones de los derechos y libertades previstos en la Carta de Derechos Fundamentales de la UE, lo cual resulta absolutamente inaceptable.

Y es que yo entiendo que legislar en el ámbito de la UE no es ni fácil ni rápido, pero desde luego su pesada maquinaria no puede ir en contra de los derechos y libertades de sus ciudadanos, puesto que es un absoluto sinsentido. Y digo esto cuando todavía, a día de hoy, no se ha aprobado el texto definitivo de Reglamento europeo sobre IA (y no se prevé que se publique hasta el año 2022), lo cual continúa provocando una inseguridad jurídica y una desprotección tremenda para los ciudadanos, puesto que a pesar de que se dispone de un Libro Blanco sobre IA, ello no es suficiente, y el embrollo y el desorden jurídico actual están pasando factura.

España, por su parte, aun ostentando soberanía para hacerlo, tampoco se ha decidido a regular el uso de la IA, a pesar de que contaba con la guía de principios básicos establecidos en el mencionado Libro Blanco europeo sobre IA, si bien ahora entiendo que ya está esperando a la publicación del antedicho Reglamento, que será directamente aplicable a nivel nacional. Y es que nuestro país, por desgracia, no presentó ni siquiera una Estrategia Nacional de IA hasta

2021, a pesar de que países de todo el mundo empezaron ya a presentarlas en 2017 (el primero fue Canadá), lo cual resulta absolutamente injustificado y muy decepcionante, sobre todo porque en noviembre de 2017 se anunció por parte del Gobierno de España la constitución de un Grupo de Sabios sobre IA y *Big Data*[800] para la publicación de un Libro Blanco español sobre IA que, no obstante, nunca llegó a ver la luz.

A pesar de lo expuesto, sin embargo, como ya he dicho, no pierdo la esperanza. Y, desde luego, una de las razones por las que decidí realizar esta tesis doctoral fue justamente la de estudiar a fondo las posibilidades que brinda la IA en materia de investigación criminal para poder así resultar de utilidad, eventualmente, a la UE o a nuestro país, en caso de que se decidan por fin a establecer una regulación clara y específica en tal sentido, ya que sé que el campo de la investigación penal a través de la IA, por desgracia, no está muy explorado puesto que resulta poco rentable. Y es que bien es sabido que donde se lleva a cabo una mayor inversión y un mayor nivel de investigación tecnológica es en el sector privado, no en la Administración, que lamentablemente no suele destinar grandes partidas presupuestarias a tales asuntos (y, si lo hace, por lo general luego no suelen conseguirse resultados eficientes y exitosos).

Tal y como se ha ido exponiendo a lo largo del presente trabajo, las utilidades de las herramientas de IA para investigar delitos (tanto herramientas de predicción y evaluación de riesgos, como herramientas de investigación criminal propiamente dichas, y herramientas de tramitación) conllevan, desde luego, una revolución sin precedentes en el ámbito de la instrucción penal.

Y es que, el uso de sistemas capaces de bucear entre millones de datos por segundo y auxiliar al juez en la toma de decisiones de forma cautelar; el empleo de sistemas que permiten identificar o verificar la identidad de una persona a través del análisis de sus datos biométricos; la utilización de sistemas capaces de interactuar con las víctimas o testigos y ejecutar comandos hablados, así como de filtrar y organizar la información contenida en millones de documentos o de detectar denuncias falsas; el uso de herramientas capaces de hacer "ciberpatrullaje", o de traducir de forma simultánea y transcribir las conversaciones

[800] Gobierno de España, 2017.

telefónicas intervenidas o las declaraciones judiciales; la utilización de herramientas con potencial para analizar imágenes, leer matrículas a tiempo real o detectar documentos falsos; o el empleo de herramientas con capacidad para descubrir fraudes y estafas y detectar la producción de disparos son, sin duda, fenómenos que pueden incrementar hasta niveles insospechados la eficiencia de la investigación criminal.

Sin embargo, como también se ha dicho, tales beneficios deben ponerse en equilibrio con los potenciales riesgos que los mencionados sistemas pueden provocar, siendo que además las posibilidades de que los resultados de estos ejerzan una influencia real en la toma de decisiones policiales, fiscales y judiciales son cada vez más elevadas.

Y es que la solución a tal cuestión, no obstante, bajo mi punto de vista, tal y como ya he venido anunciando de forma reiterada, está muy clara: la regulación (sin caer en el exceso) y, sobre todo, la creación de mecanismos que garanticen su estricto cumplimiento.

En mi opinión, pues, debe ser el poder legislativo el que, en cada momento decida cómo quiere llevar a cabo el mencionado equilibrio entre las posibles utilidades y los potenciales riesgos que conlleva el uso de las herramientas de IA de investigación de delitos (con sujeción y respeto, en su caso, a los principios básicos comunes que pudieran llegar a aprobarse a nivel internacional), especialmente atendiendo a su posicionamiento político, en concreto, respecto de la ecuación libertad-seguridad.

Para ello, no obstante, considero que el poder legislativo (tanto europeo como nacional) debe estar asesorado de forma transversal por los más prestigiosos y brillantes expertos en materia de IA, de protección de datos personales, de ciberseguridad y de Derecho Penal, puesto que legislar sobre una materia tan compleja y peligrosa como la del uso de tal tecnología en el ámbito de la investigación penal, con la cantidad de derechos fundamentales que hay en juego, debe llevarse a cabo de la forma más garantista y cautelosa.

Y es que bien es sabido que no todo lo técnicamente posible tiene por qué ser jurídicamente viable, por lo que resulta, bajo mi punto de vista, absolutamente necesaria una colaboración multidisciplinar con capacidad de auxiliar al poder legislativo de la forma más técnica, objetiva y acertada posible en su tarea de legislar sin perjuicio de que luego este, en función

de su tendencia de cada momento, opte por tomar las decisiones legislativas que considere más convenientes, más o menos garantistas, o más o menos flexibles.

En cualquier caso, lo que sí que considero fundamental es que las leyes que regulen el uso de la IA para la investigación criminal sean específicas, completas y, sobre todo, claras. Y pongo énfasis en este último requisito porque tal legislación deberá servir como guía no solo para los operadores jurídicos que la apliquen, sino también para los diseñadores y desarrolladores de sistemas que, desde luego, deberán crearlos para que puedan cumplir con las exigencias legales. Y eso es lo que, en conversaciones mantenidas con diversos ingenieros para realizar el presente trabajo, se me ha trasladado de forma más común y vehemente, puesto que según lo que me exponen, les resulta muy complicado trabajar sin una guía jurídica clara de actuación y, como consecuencia de ello, en ocasiones programan sistemas que luego son señalados como ilegales y resultan demonizados (y, no me extraña, porque el vacío jurídico y el guirigay legislativo que hay hoy en día "clama al cielo").

QUINTA.- Los actuales pasos europeos y españoles hacia la regulación del uso de la IA en el ámbito de la investigación penal son una buena noticia pero resultan tardíos y, en ocasiones, insuficientes.

A pesar de que en el ámbito de la UE y de España, como ya he advertido con anterioridad, actualmente por desgracia no contamos con una regulación clara y específica sobre IA, procede mantener la atención, por un lado, en el que, al parecer, como ya se ha dicho, va a devenir el texto del futuro Reglamento sobre IA que, no obstante, no es todavía definitivo, ya que por el momento únicamente se cuenta con la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión; y, por otro lado, en el nuevo texto de la Ley de Enjuiciamiento Criminal española que está en ciernes, aunque por ahora no hay noticias de que vaya a incluir una regulación específica sobre el uso de los sistemas de IA en el ámbito de la investigación penal, lo cual por un lado es una mala noticia pero por otro lado se entiende prudente, habida cuenta de que se está a la espera de la publicación del mencionado Reglamento europeo que, sin duda, tendrá implicaciones en tal materia.

Dicho esto, tal y como ya he anunciado en páginas anteriores, entiendo que la regulación de la IA, en concreto en el ámbito de la investigación penal, que es el que nos ocupa, para resultar de utilidad, debe, por un lado, definir y limitar muy bien en qué casos y en qué circunstancias pueden ser empleados los sistemas de IA; por otro lado, establecer mecanismos para garantizar la calidad de los mismos tanto en el momento de su comercialización o puesta en circulación como con posterioridad; y, finalmente, disponer mecanismos para garantizar el cumplimiento de los requisitos legales vigentes.

Respecto del primero de los aspectos a regular (a saber, los casos y los términos en que pueden ser empleados los sistemas de IA en el ámbito de la investigación penal), es interesante poner de manifiesto que la ya mencionada Propuesta de Reglamento sobre IA, por un lado, establece qué sistemas están prohibidos y cuáles pueden emplearse en cada caso; y, por otro lado, determina bajo qué circunstancias debe llevarse a cabo, en su caso, su uso.

Así, por ejemplo, tal y como ya se expuso en la Sección 3.2.2.2. (A.3), en el artículo 5.1.d) se prevé la prohibición a las autoridades de la utilización de sistemas de identificación biométrica remota en tiempo real en espacios de acceso público con fines de aplicación de la ley, salvo que su uso sea estrictamente necesario para:

-la búsqueda selectiva de posibles víctimas concretas de un delito, incluidos menores desaparecidos;
-la prevención de una amenaza específica, importante e inminente para la vida o la seguridad física de las personas físicas o de un atentado terrorista; o
-la detección, la localización, la identificación o el enjuiciamiento de la persona que ha cometido o se sospecha que ha cometido alguno de los delitos mencionados en el artículo 2.2 de la Decisión Marco 2002/584/JAI del Consejo, siempre que la normativa en vigor en el Estado miembro implicado imponga una pena o una medida de seguridad privativas de libertad cuya duración máxima sea de, al menos, tres años.

Y es que, como en este caso, a lo largo de la mencionada Propuesta se disponen una serie de limitaciones relativas al uso de herramientas de IA en el ámbito de la investigación penal que resultan de enorme utilidad, tanto para los creadores y diseñadores de sistemas, como para los

operadores jurídicos y autoridades, que por fin, si llega a aprobarse el Reglamento sobre IA proyectado, tendrán una guía clara para actuar.

Y, respecto de todas aquellas herramientas que no constan prohibidas, se prevén en la Propuesta una serie de circunstancias legitimadoras para su uso y una serie de requisitos de obligado cumplimiento, lo cual es también de enorme utilidad.

En relación con ello, y siguiendo con el caso anterior, se prevé, por ejemplo, que el uso de los sistemas de identificación biométrica remota en tiempo real en espacios de acceso público, en aquellos casos en que resulten legalmente permitidos por ser estrictamente necesarios para conseguir alguno de los antedichos fines, deba ser autorizado por un juez o por una autoridad administrativa independiente del Estado miembro donde vaya a llevarse a cabo. Y, además, se dispone que tal autorización deberá basarse en lo previsto legalmente, en su caso, por la normativa nacional de cada Estado miembro, en los términos previstos en el propio Reglamento.

Y es que el uso de tal clase de sistemas, en concreto, implica el riesgo de que nuestro día a día se convierta en una especie de "Gran Hermano" que suponga la pérdida absoluta de, principalmente, nuestros derechos a la libertad y a la intimidad, por lo que sin duda debe limitarse, ya que en contra de lo que se piensa en algunos países (China, sin ir más lejos): ¿acaso no nos echaríamos las manos a la cabeza si por ley se permitiera que un policía procediera a seguirnos cada vez que saliéramos de casa y fuera testigo de todos nuestros movimientos? Desde luego que sí, y ello, sin duda, es exactamente lo mismo que tener cámaras colocadas en la vía pública que rastreen todos nuestros movimientos a través de la identificación biométrica remota en tiempo real y, por tal razón, desde la UE con acierto se pretende limitar su uso.

Todo ello, no obstante, debe ponerse en relación con lo previsto actualmente en la legislación en materia de protección de datos personales, que si bien no regula de forma directa y específica el uso de los sistemas de IA, sí que tiene implicaciones al respeto, habida cuenta de que estos llevan a cabo, en la mayoría de ocasiones, el tratamiento de categorías especiales de datos. Y, en concreto, procede poner el foco en lo previsto en la LO 7/21, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y

enjuiciamiento de infracciones penales y de ejecución de sanciones penales, especialmente en sus artículos 13 y 14, tal y como se ha venido poniendo de manifiesto en la presente tesis doctoral.

Y es que los sistemas de IA, como ya se ha dicho, esencialmente se nutren de datos, habiendo sido estos catalogados como "el petróleo del siglo XXI", por el enorme valor que han alcanzado, ya que se usan para entrenar a tales herramientas con el fin de que estas mejoren su calidad.

No obstante, "el diablo está en los datos", ya que como es ya sabido no vale cualquier tipo de información para conseguir tal fin, puesto que los datos que nutren a un sistema de IA deben ser de calidad, ya que en caso contrario, lo único que se consigue es, por un lado, perpetuar patrones históricos en muchos casos ilegales y, por otro lado, crear nuevas situaciones antijurídicas que vulneren derechos y libertades de los ciudadanos.

Y es que, desde luego, como se ha ido viendo, la calidad de los sistemas de IA depende de sus características técnicas pero, sobre todo, de los datos que emplean, ya que son el elemento clave de tal tecnología.

Así, si una herramienta de IA usa datos poco representativos, sus niveles de precisión tienden a disminuir y los riesgos de vulneración del derecho a la igualdad y a la no discriminación, a aumentar; y si una herramienta de IA trata datos obtenidos de forma ilegal o lleva a cabo tratamientos que no cumplen con los requisitos legalmente previstos, el riesgo de vulneración del derecho a la protección de datos personales, entre otros, tiende a subir, por lo que resulta fundamental, bajo mi punto de vista, poner especial atención al tipo de datos que se emplean y, sobre todo, al tratamiento que se les da.

Y, desde luego, lo expuesto hace que resulte absolutamente necesario que quede claro, *ex lege*, en qué casos y bajo que circunstancias puede llevarse a cabo el tratamiento de los datos personales para el uso de herramientas de IA en la instrucción de las causas, habida cuenta de que un error en tal sentido podría tener nefastas consecuencias. Y es que, entre otras cuestiones, ¿qué ocurriría si una herramienta de IA empleada por un juez, luego, ante la impugnación efectuada por la defensa, se decreta ilegalmente utilizada por indebido

tratamiento de los datos personales, por ejemplo? Desde luego, ello conllevaría la nulidad de todas las decisiones tomadas con base en los resultados por ella arrojados (con riesgo de causación de daños irreparables), pero, sin duda, al juez le resultaría muy complejo abstraerse de los mismos, al igual que ocurriría, por ejemplo, con el contenido de una prueba ilícita que no puede valorarse oficialmente pero se conoce de forma extraoficial. Un auténtico despropósito.

Respecto del segundo de los aspectos a regular (a saber, los mecanismos para garantizar la calidad de los sistemas de IA de investigación criminal, tanto en el momento de su comercialización o puesta en circulación, como con posterioridad), es interesante poner de manifiesto que la ya analizada Propuesta de Reglamento sobre IA establece una serie de procedimientos, requisitos y controles que varían atendiendo, especialmente, a su calificación o no como sistemas de IA de alto riesgo según lo dispuesto en su artículo 6.

Y es que está claro que deben adoptarse todas las cautelas posibles para conseguir que los sistemas de IA se empleen en beneficio de la investigación sin vulnerar ningún derecho y, sobre todo, sin generar situaciones antijurídicas que pueden resultar irreversibles.

En relación con ello, como se ha dicho, la antedicha Propuesta fija unos procedimientos de evaluación de la conformidad con la legalidad vigente de los sistemas de IA calificados de alto riesgo[801], lo cual celebro. No obstante, bajo mi punto de vista el modelo propuesto no es el más acertado e idóneo (por insuficiente) para garantizar la existencia de un verdadero filtro y la concurrencia de una real verificación institucional de la calidad de las antedichas herramientas. Me explico.

Entiendo que el sistema de evaluación de conformidad propuesto en el mencionado texto es quizás el más viable económicamente y el que más encaja con los actuales medios y la presente configuración de las instituciones europeas, pero tal y como he venido advirtiendo, considero que la IA es una materia que entraña graves peligros y que, por ende, debe ser

[801] Véanse págs. 192-197.

tratada con las mayores cautelas, aunque supongan una inversión mayor, máxime cuando se está haciendo referencia a sistemas calificados por la propia normativa como de alto riesgo.

Mi propuesta al respecto, que ya ha sido anunciada en varias ocasiones a lo largo de la presente tesis doctoral, no es otra que la de crear, en el ámbito de la UE, una Agencia sobre IA que tenga competencia para examinar, filtrar y verificar la calidad de los sistemas y su conformidad con la legalidad vigente, con carácter previo a su introducción en el mercado o a su puesta en circulación, y con posterioridad. Tal Agencia, a mi modo de entender, podría tener una sede central y, a su vez, delegaciones en los distintos Estados Miembros, con el fin de poder descentralizar tales operaciones y que estas se llevaran a cabo con mayor agilidad y más facilidad, siempre bajo unos procesos comunes y únicos, controlados en última instancia desde la sede principal.

Dicho modelo, similar al que sigue la Agencia Europea del Medicamento, en mi opinión es el más óptimo para llevar a cabo las tareas de análisis y verificación de la calidad y conformidad de los sistemas, ya que, por un lado, sería una institución pública la que habría detrás del proceso, lo cual es fundamental a efectos de generar confianza en los ciudadanos, puesto que el interés único es (o, al menos, debería ser) el oficialmente anunciado; y, además, hay un riesgo muy bajo de filtración de las patentes millonarias que suele haber detrás de tales herramientas.

No obstante, tal y como se ha expuesto en la Sección 3.1., en la Propuesta de Reglamento europeo sobre IA se contemplan unos procedimientos de verificación de la calidad de los sistemas distintos al que yo propongo y, solo de forma excepcional se prevé la intervención activa de las autoridades, lo cual, como he dicho, no me parece lo más oportuno.

Y, finalmente, respecto del último de los aspectos a regular (a saber, los mecanismos para garantizar el cumplimiento de los requisitos legales vigentes), procede poner de manifiesto que la analizada Propuesta de Reglamento europeo sobre IA, tal y como se ha expuesto en la Sección 3.1., establece en su artículo 71 una serie de sanciones para casos de incumplimiento y delega además en los Estados miembros la creación de un régimen de sanciones para los supuestos de vulneración de lo establecido en el futuro Reglamento, lo cual aplaudo y considero fundamental para el éxito de la legislación sobre tal tecnología.

Además de lo expuesto, la mencionada Propuesta incluye algunas medidas de gobernanza tales como la creación de un Comité Europeo de IA con finalidad de ofrecer asesoramiento y asistencia a la Comisión en su tarea de coordinar y garantizar el cumplimiento de lo dispuesto en el mencionado cuerpo legal; y, asimismo, prevé que cada Estado miembro establezca o designe autoridades nacionales competentes con el fin de garantizar la aplicación y la ejecución del futuro Reglamento, lo cual entiendo fundamental, puesto que considero que es indispensable que se vele por el cumplimiento de la legislación en materia de IA, habida cuenta de la cantidad de derechos fundamentales que podrían verse vulnerados en caso contrario y, sobre todo, habida cuenta de que la credibilidad de la UE en materia de garantías de los sistemas de tal tecnología está en juego. Y, en relación con ello, con el fin de aumentar la protección, además, de forma acertada, bajo mi punto de vista, se prevén una serie de sanciones para casos de incumplimiento de lo establecido en el futuro Reglamento.

En cualquier caso, y con el fin de establecer controles previos, considero fundamental la creación de espacios controlados de pruebas para la IA (los denominados *sandboxes*), lo cual de forma muy acertada ya está previsto en la mencionada Propuesta. Y es que tales espacios, tal y como dispone su artículo 53.1, proporcionan un entorno controlado que facilita "*el desarrollo, la prueba y la validación de sistemas innovadores de IA durante un periodo limitado antes de su introducción en el mercado o su puesta en servicio, en virtud de un plan específico*", lo cual es altamente beneficioso, especialmente si se lleva a cabo bajo la supervisión y la orientación directas de las autoridades competentes, como se prevé en dicho texto.

SEXTA.- En todo caso, el uso de la IA por parte de las autoridades en el proceso penal de instrucción debe ser llevado a cabo de forma responsable y cautelosa, con especial atención al principio de transparencia.

Y es que, debo poner de manifiesto que, incluso en caso de que el uso de los sistemas de IA de investigación criminal quede legalmente limitado y estos estén sujetos a controles de calidad, entiendo que, habida cuenta de los enormes riesgos que, *per se*, estos entrañan, las autoridades que los utilicen deberán hacerlo con la mayor diligencia y cautela posibles, al igual que hacen los médicos cuando recetan medicamentos o los farmacéuticos cuando los

dispensan, a pesar de que estos se hallen autorizados para unos fines concretos por la Agencia Europea del Medicamento.

Así, en especial, entiendo que los policías, los fiscales y los jueces que se hallen legitimados para aplicar sistemas de IA con fines de investigación delictiva, deberán llevar a cabo un uso muy escrupuloso y riguroso de los mismos, con pleno respeto y garantía en todo caso y bajo cualquier circunstancia al derecho a la defensa, para lo cual las personas que se vean afectadas deberán tener conocimiento constante y absoluto del contenido de las herramientas aplicadas, a los efectos de poder entenderlo y, en su caso, cuestionarlo y rebatirlo, en aplicación del principio de contradicción, tanto durante la fase de instrucción como en el acto del plenario. Y ello, sin duda, requiere que los sistemas de IA que se empleen en el ámbito de la investigación criminal cumplan con los más estrictos requisitos de transparencia, puesto que es la única forma de poder determinar qué hay detrás de los resultados que sirven para tomar aquellas decisiones que afectan a las personas implicadas.

No obstante, en relación con esto último, respecto de la gran polémica que, bajo mi punto de vista, con razón, se ha generado en torno a la necesidad de transparencia de los sistemas de IA, en especial en casos de uso en el ámbito de la investigación criminal, me gustaría reiterar una reflexión ya realizada en el cuerpo de la presente tesis doctoral que me parece importante.

Y es que, si bien, desafortunadamente, en la actualidad, la mayoría de veces la toma de decisiones con algoritmos deviene sinónimo de toma de decisiones inexplicable (lo que suele ser definido como caja negra o *black box*), y ello es inaceptable, no debemos olvidar que los policías, los fiscales y los jueces, como humanos que son, también cuentan con su particular caja negra o *black box* (a saber, su cerebro) que, de hecho, es mucho peor que la de los sistemas de IA, puesto que deviene en todo caso inescrutable.

Respecto de ello entiendo que, si bien existen diferencias patentes entre las potenciales consecuencias que puede tener la toma de decisiones efectuada por un humano y la realizada por una máquina -especialmente por el alcance que puede llegar a lograr, que se multiplica de forma exponencial en caso de resultar automatizada-, lo cierto es que ambas cuentan con elementos comunes que, sin duda, deben ser tenidos en cuenta y valorados, principalmente con el fin de establecer mejoras, puesto que de ello va la evolución.

Así, desde luego, no por el mero hecho de que los humanos cuenten con su particular caja negra o *black box* debe permitirse que los sistemas de IA carezcan de transparencia (y menos en el ámbito de la Administración Pública), puesto que, entre otras cosas, las autoridades que toman decisiones en el ámbito de la investigación penal cuentan siempre con la obligación legal de fundamentarlas jurídicamente (al menos, formalmente, sin perjuicio de las reales motivaciones que les hayan llevado a resolver un asunto de un modo determinado, que son casi imposibles de descubrir), lo cual no sucede con las máquinas.

De acuerdo con ello, considero que la IA, en tal sentido, a pesar de las actuales (y fundadas) dudas al respecto, puede llegar a suponer un avance sin precedentes en el ámbito de la transparencia en la justicia, habida cuenta de que, si se logra fijar un filtro real que solo permita el uso de sistemas de calidad, con garantía de transparencia, explicabilidad y trazabilidad, se conseguirá llevar tal requisito tan fundamental al máximo nivel, con una clara mejora de los estándares actuales.

Y ello resulta especialmente importante en un momento en que está en absoluto auge la idea del denominado "juez robot", que tiene como principal finalidad la de automatizar la resolución de litigios *on line* a través de la IA.

A modo de ejemplo, entre otros, por un lado, en 2017 se inauguró en Hangzhou (China) el primer "*Internet Court*" para resolver conflictos relacionados con cuestiones comerciales (que hoy en día cuenta con un 99% de aceptación de las sentencias y su duración media de celebración de juicios y resolución de casos es de veintiocho minutos).[802] Desde entonces, se han instaurado tribunales de tal clase, asimismo, en Pequín y Guangzhou y, entre todos ellos, según lo dispuesto en el informe emitido en diciembre del 2019 por el Tribunal Supremo chino, han resuelto más de cien mil casos desde su puesta en funcionamiento[803]. Además, en marzo del 2019, en doce provincias de China se lanzó un "*Mobile Court*" a través de la

[802] Véase Hangzhou Internet Court, s.f..
[803] Tribunal Supremo chino, 2019, pág. 64.

plataforma WeChat[804], y según manifestó el mencionado Alto Tribunal chino en el antedicho informe, desde su lanzamiento este ha resuelto tres millones de casos.[805]

Por otro lado, en Estonia, ya en agosto del 2019 el gobierno encomendó a su Director de Datos, Ott Velsberg, la coordinación y supervisión de la creación de un sistema de IA que permitiera la resolución *on line* de disputas menores, inferiores a siete mil euros, que hoy en día está ya en pruebas piloto para entrar en funcionamiento.[806]

SÉPTIMA.- De entre todos los riesgos inherentes al uso de la IA en el ámbito de la instrucción penal hay uno especialmente peligroso: el de la deshumanización de la justicia.

En particular, no obstante, y tras hacer referencia a los potenciales riesgos de los sistemas de IA y su posible solución (a saber, la regulación), debo decir que de todos los peligros expuestos, hay uno que, por encima del resto, me preocupa especialmente: el de la deshumanización de la justicia.

Y es que, la acepción de justicia que más se ajusta a lo que yo entiendo como tal es aquella que la define como: "*Principio moral que lleva a dar a cada uno lo que le corresponde o pertenece*."[807] . No obstante, tal y como se expuso en la Sección 2.4.1., bien es sabido que la moral es un concepto abstracto e indeterminado, cuyo contenido varía en función de las creencias y valores de cada persona.

Y, ¿qué hay de especial en tal aseveración? Pues justamente la singularidad de la concepción de moral y, por ende, de justicia, que tiene cada uno de los individuos que habitan el mundo, no existiendo, a mi entender, dos idénticas. Y es que puede que estas sean parecidas, similares o aparentemente iguales, pero nunca idénticas. Y resulta evidente que los policías, los fiscales y los jueces son seres humanos, y como tales, asimismo, tienen sus propios conceptos de justicia.

[804] Véase Wang, 2019.
[805] Tribunal Supremo chino, 2019.
[806] Véase The Technolawgist, 2019.
[807] Diccionario de la RAE.

Está claro que, hoy en día, España es un Estado de Derecho en que la ley es la que prevalece sobre cualquier tipo de opinión o valoración personal a la hora de impartir justicia. No obstante, lo cierto es que en muchas ocasiones la ley otorga la posibilidad a las autoridades que la aplican de llevar a cabo interpretaciones, valoraciones de circunstancias o realización de excepciones basándose en su propia experiencia o en las particularidades del caso concreto. Sí, en la justicia tiene cabida la flexibilidad. Y esta será mayor o menor en función del asunto, en función de lo que dispongan la ley aplicable y la jurisprudencia y, por qué no, en función del concepto de justicia que tenga la autoridad competente que la aplique en cada momento. Y es que sería absurdo negar que existen policías, fiscales y jueces más y menos rígidos, más y menos experimentados, más y menos empáticos, y más y menos voluntariosos, puesto que ello es una realidad.

Así, por mi experiencia como juez de instrucción puedo decir que sé a ciencia cierta que algunos de los casos que han pasado por mí hubieran sido resueltos de cientos de formas distintas por mis compañeros. Y todos ellos hubieran actuado con pleno respeto a la ley, pero cada uno con inclinación hacia su propio concepto de justicia, con el objetivo siempre de ser lo más equitativos y objetivos posible. Y es que en ocasiones, la ley da facultad para ello. Así, la parte más humana de la justicia, sin duda, existe.

¿Cuántas veces los agentes de policía y los fiscales tratan de interactuar con las partes implicadas en un caso y/o sus Letrados para que intenten solucionar sus problemas, poniendo solución al fondo, si es posible, y eviten la vía penal para ahorrarse mayores perjuicios? Yo, desde luego, en ciertos casos, por ejemplo, remito a las partes a mediación y/o a tratamiento psicológico en caso de ver que ello podría resultar beneficioso para los interesados, y en ocasiones tiene efecto. ¿Cuántas veces se da una mirada de cariño y comprensión a una víctima o se le ofrecen palabras de aliento? ¿Cuántas veces se intenta poner remedio médico a personas que están metidas en verdaderos bucles delictivos por sus problemas de salud mental o de adicción a sustancias estupefacientes? ¿Cuántas veces se advierte a un individuo de que, de seguir así, va a acabar en prisión y se le intenta motivar para que corte con su vida anterior y empiece una nueva etapa? ¿Cuántas veces se levanta el teléfono para hablar directamente con los Servicios Sociales en casos de riesgo y se trata de solucionar asuntos que requieren de conversaciones entre humanos? Todo ello ocurre, porque somos personas, y, no

obstante, resulta absolutamente incompatible con una toma de decisiones automatizada efectuada por una máquina, basada únicamente en los algoritmos.

Y es que, por un lado, procede decir que el despropósito y el perjuicio emocional que puede suponer para una persona que se halla en una situación de máxima vulnerabilidad (como ocurre en el prácticamente 100% de los casos en el ámbito de la investigación penal) interactuar con una máquina puede ser enorme, habida cuenta de que ello genera una impotencia inmensa, puesto que ya de entrada impide la valoración, por parte de un humano con cierta empatía, de matices que, por ahora, solo una persona puede captar (y la vida, justamente, va de matices); y, por otro lado, debe ponerse de manifiesto que el respeto pleno del derecho a la dignidad de la persona, que es la esencia pura del ser humano y lo que nos define como especie, podría verse resquebrajado en caso de eliminar la intervención humana, lo cual, bajo mi punto de vista, resulta absolutamente inadmisible. Y no solo porque nos desnaturaliza como especie sino porque, además, de verdad creo que no aportaría, ni muchísimo menos, en términos generales, mejores resultados que los que arroja el sistema actual.

Por suerte, no obstante, en la mayor parte de países del mundo, tal y como ya se ha ido poniendo de manifiesto en el presente trabajo, la idea de que la IA debe estar centrada en el interés del ser humano, no debe sustituir íntegramente las relaciones personales y debe ser siempre controlada y supervisada por una persona, está fuertemente arraigada. Y en tal sentido, afortunadamente, van todos los proyectos, iniciativas y regulaciones a los que se ha hecho referencia en esta tesis doctoral, puesto que la humanidad sabe que tras el uso de la IA hay mucho en juego. Y con base en tal premisa, bajo mi punto de vista, debe enfocarse el uso de tal tecnología en el ámbito de la investigación penal, erigiéndose como complemento idóneo para las autoridades, sin llegar a sustituirlas.

Así, a modo de ejemplo, el artículo 14 de la LO 7/21, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, ya mencionado con anterioridad, prohíbe con carácter general las decisiones basadas únicamente en un tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos negativos para el interesado o que le afecten significativamente, salvo que se autorice expresamente por

una norma con rango de ley o por el Derecho de la Unión Europea. Y a ello se añade, además, que la norma habilitante del tratamiento "*deberá establecer las medidas adecuadas para salvaguardar los derechos y libertades del interesado, incluyendo el derecho a obtener la intervención humana en el proceso de revisión de la decisión adoptada*.", lo cual sin duda va en la línea del respeto a los valores expuestos.

Sin perjuicio de ello, la pandemia del COVID-19 ha demostrado al mundo de forma más clara que nunca que la inversión y la apuesta por la tecnología y, en concreto, por la IA, puede ser el perfecto aliado para los seres humanos en caso de resultar empleada de forma correcta. Y es que como se ha venido diciendo, las ventajas del uso de sistemas de IA en la mayor parte de los ámbitos de la vida, son enormes. Lo que no puede perderse de vista, sin embargo, aunque resulte a veces tentador, es que su diseño, desarrollo y utilización deben respetar, defender y promover los valores y los derechos y libertades de los ciudadanos de las sociedades donde estos operan, tratando que ello, no obstante, no suponga un impedimento para la prosperidad y la competitividad (equilibrio que, sin duda, es difícil de conseguir)

OCTAVA.- Conclusión final: la IA puede resultar muy útil para mejorar la eficiencia del proceso penal de instrucción español <u>únicamente</u> si su uso está legalmente previsto. Breve reflexión sobre el valor probatorio de las diligencias practicadas con herramientas de IA.

En relación con todo lo expuesto, pues, y a modo de conclusión final, tras haber llevado a cabo la investigación que consta en el cuerpo de la presente tesis doctoral, me veo en posición de responder a la siguiente pregunta, configurada como la hipótesis que ha dado razón a este trabajo: ¿puede, entonces, resultar útil la IA para investigar delitos en el proceso penal de instrucción español? La respuesta, claramente, es "sí", pero con matices. Matices que, no obstante, tal y como he ido exponiendo, se traducen principalmente en potenciales riesgos que pueden (y deben) ser mitigados mediante la creación de unos principios básicos comunes de actuación, a nivel internacional, y la elaboración de leyes nacionales (o transnacionales, como en el caso de la UE) claras, específicas y completas.

Y ante ello, inevitablemente, surge la duda de cuál sería el valor jurídico, especialmente probatorio, que podría llegar a otorgarse hoy en día en España a las herramientas de IA

empleadas por las autoridades en el proceso de instrucción penal, lo cual, empero, podría dar lugar a la elaboración de una completa tesis doctoral independiente y excede del ámbito de la presente. No obstante, considero importante hacer una breve reflexión al respecto.

Y es que, tal y como se ha advertido de forma reiterada a lo largo del presente trabajo, en nuestro país actualmente carecemos de una regulación especial sobre el uso de la IA por parte de las autoridades en el ámbito de la investigación penal. No obstante, sí contamos con legislación sobre protección de datos personales, que como se ha ido exponiendo contiene algunas referencias que son de aplicación a tales sistemas, y también contamos con una Carta Magna y una Carta de Derechos Fundamentales de la UE que no solo son claras sino que además han sido vastamente interpretadas por la jurisprudencia. Y tales instrumentos legales son los que, desde luego, entiendo que deberían operar hoy en día como guía para, por un lado, llevar a cabo el uso y la aplicación de los sistemas de IA por parte de las autoridades en el ámbito del proceso de instrucción penal y, por otro lado, determinar su valor. Me explico.

A falta de leyes que de forma concreta ofrezcan pautas a los operadores jurídicos y a las autoridades (incluidas las fiscales y judiciales) para emplear las mencionadas herramientas en la investigación criminal, procede acudir a nuestros principios básicos y a las posibles leyes relativas a otras materias que resulten aplicables al caso concreto. Y es que siendo que en la actualidad, por desgracia, todavía no contamos con un sistema público de certificación de calidad de los sistemas de IA que permita emplearlos con todas las garantías, debemos tomar todas las cautelas para evitar llevar a cabo vulneraciones de derechos fundamentales.

Por un lado, así, por ejemplo, en caso de que en un proceso penal la policía presentara al juez un atestado en que se pusiera de manifiesto el uso de una herramienta de IA de "ciberpatrullaje" que hubiera detectado la presencia en la red de varios presuntos terroristas dispuestos a atentar, ello no debería tenerse por cierto y correcto por el instructor de la causa de forma automática puesto que, como es sabido, detrás de tal clase de *software* hay muchos riesgos latentes. No obstante, tampoco se reputaría prudente desechar de antemano la información otorgada por el sistema, habida cuenta de que también se ha demostrado que en ocasiones funcionan.

De acuerdo con ello, considero que en tales supuestos lo que debería hacerse es una ponderación de derechos para poder determinar, en cada caso concreto, el valor que procedería otorgar a los resultados presentados por la herramienta de IA empleada.

Para ello, principalmente, entiendo que debería atenderse al origen y al contenido del sistema de IA utilizado, con obligación de averiguar qué compañías hay detrás de su diseño y de su comercialización, qué garantías ofrece, qué filtros de la Administración ha pasado para ser usado, qué niveles de precisión maneja y, sobre todo, con qué grado de transparencia y explicabilidad cuenta.

Yes que, por desgracia, hoy en día es muy difícil que un sistema de IA (con alta probabilidad creado y distribuido por alguna empresa de *software* extranjera y adquirido por autoridades españolas) pueda tener respuestas satisfactorias a todas esas preguntas.

Así, visto lo visto, considero que el juez instructor, en tal caso, debería requerir, sin excepción, que la herramienta de IA empleada por la policía fuera transparente y explicable, con el fin de poder analizar, entender y contrastar su contenido. Además, no obstante, entiendo que el juez debería exigir un *plus* de indicios para poder tener en cuenta la información remitida por la policía, que debería recopilar, con base en los datos obtenidos por el sistema de IA, los máximos datos posibles para incriminar a las personas sospechosas con la finalidad de otorgar al juez la mayor información posible para que este pudiera otorgar a los resultados presentados, al menos, el valor de *notitia criminis*.

Y es que, desde luego, cualquier decisión judicial basada únicamente en el resultado de un *software* opaco resultaría nula de pleno derecho (incluso aunque luego se llegara a demostrar que tenía razón), puesto que ello devendría contrario, en todo caso, al derecho a un proceso con todas las garantías y al derecho a la presunción de inocencia, sin perjuicio de las ulteriores vulneraciones de derechos que su uso pudiera entrañar (entre otras, derecho a la defensa, derecho a la intimidad, derecho a la libertad, etc). Y potencialmente nula resultaría también, por los derechos que podría conculcar, cualquier tipo de decisión judical basada solamente en el resultado de un sistema de IA transparente cuya calidad no constara públicamente certificada, habida cuenta de su posible falta de precisión, de la posible mala calidad de sus

datos y de la consiguiente existencia de sesgos, de la potencial presencia de brechas de seguridad y, en su caso, de la dificultad de rendición de cuentas.

Y, asimismo, en caso de que se empleara por el juez instructor un sistema de reconocimiento facial, por ejemplo, para identificar al autor de un delito investigado, entiendo que aplicarían los mismos estándares.

Y es que la *conditio sine qua non* para que, al menos, el resultado arrojado por una herramienta de IA pudiera contar con valor de *notitia criminis* e, incluso, diligencia de investigación o indicio, considero que sería su transparencia y explicabilidad, ya que solo con tales características las partes personadas, el Ministerio Fiscal y el juez podrían tener conocimiento real del contenido de los algoritmos para, en su caso, someterlos a contradicción (con la finalidad de detectar la posible existencia de fallos, errores, datos ilegales, brechas de seguridad y otras cuestiones que podrían conllevar la nulidad o pérdida de valor.

No obstante, todo lo expuesto se verá modificado, con alta probabilidad, cuando se apruebe por fin el Reglamento europeo sobre IA, que espero establezca pautas claras no solo para el uso de los sistemas de IA por parte de las autoridades en el ámbito del proceso penal sino también para determinar su valor jurídico y, sobre todo, fije por fin un sistema público de certificación de calidad que permita a las autoridades emplear herramientas de IA sin un miedo insuperable de cusar lesiones de derechos. Y todo ello me motiva, sin duda, a llevar a cabo una investigación postdoctoral en el futuro.

Y es que estoy segura de que, una vez se establezca un marco jurídico claro que delimite el uso de la IA en el ámbito de la investigación criminal y garantice unos estándares de calidad elevados, tanto las autoridades, como los operadores jurídicos y los ciudadanos irán aumentando su confianza en tal nueva tecnología, y el empleo de esta se irá generalizando hasta terminar convirtiéndose en un elemento habitual más de apoyo para policías, fiscales y jueces.

En tal sentido, Richard Berk, el profesor emérito de criminología de la Universidad de Pensilvania (Filadelfia, EEUU) que diseñó el algoritmo utilizado por el Departamento de Libertad Condicional del estado de Filadelfia, ya predijo que las controversias sobre el uso de

los sistemas de IA se desvanecería a medida que los algoritmos se generalizaran, y los comparó con los sistemas de piloto automático que hoy en día utilizan los aviones comerciales, habiendo manifestado al periódico The New York Times: "*Hemos aprendido que el piloto automático es confiable, más confiable que un piloto humano individual. Aquí va a pasar lo mismo*".[808]

Y es que, en mi opinión, en relación con lo que manifiesta el mencionado profesor, lo mismo que hoy en día está ocurriendo con la IA, ya ocurrió en su día con todas y cada una de las nuevas tecnologías que han ido surgiendo, habiendo sido, de hecho, calificada la IA como la nueva electricidad.[809] Y, así ocurrió, por ejemplo, con la luz, con los aviones, con los coches, con el teléfono, con la televisión, con los ordenadores y con la calculadora, inventos que en su día se veían casi como diabólicos y que hoy en día de modo natural, debidamente regulados, forman parte de nuestras vidas cotidianas.

Así, debemos tomarnos nuestro tiempo para asimilar, entender y, sobre todo, crear los mecanismos adecuados para garantizar el uso de la IA de la forma más segura y confiable posible, pero sin excedernos.

Y, en relación con ello, procede traer a colación lo manifestado por el sabio escritor Gabriel García Márquez, que aseveró: "Desde la aparición de vida visible en la Tierra debieron transcurrir 380 millones de años para que una mariposa aprendiera a volar, otros 180 millones de años para fabricar una rosa sin otro compromiso que el de ser hermosa, y cuatro eras geológicas para que los seres humanos fueran capaces de cantar mejor que los pájaros y morirse de amor."[810]

Apasionante, sin duda, la nueva era que tenemos por delante.

[808] Véase Metz & Satariano, 2020.
[809] Véase Shana, 2017.
[810] García, 1986.

4 bis- CONCLUSIONS

As already announced in the Introduction and as has been reiterated throughout this paper, the main purpose of this work was to investigate the possible uses of AI in the Spanish investigation process in order to improve its efficiency and, therefore, the quality of the service provided by the Administration of Justice to the citizen.

To this end, it has been essential not only to carry out an initial study on the concept of AI and its various applications, but also to carry out an in-depth analysis of the relationship between such technology and the Law, given the enormous risks that it entails and which, in any case, must be mitigated from the legal sphere. And, having laid the aforementioned foundations, I have proceeded to carry out an exhaustive examination of all those AI tools that could be useful in the criminal investigation of cases without, obviously, obviating their potential risks, making references in each case to the current and/or future regulations.

In any case, I feel it necessary to warn that this doctoral thesis is far from being a port of arrival, but rather a port of departure, given that the issues addressed in it are very new and will undoubtedly be the subject of many subsequent and sometimes divergent studies (especially when the European Regulation on AI is finally published). However, this seems to me to be a desirable and "healthy" scenario, as it will be the best indicator that criminal investigation is evolving, and that is essentially the spirit that, as a public servant, I would like to instill: to look forward with a will to improve. Always.

Having stated the above, it should be made clear that this work has led me to draw the conclusions which, as a personal proposal (knowing that there may be other points of view), I will now set out.

FIRST.- AI can bring great benefits and utilities to the Spanish criminal investigation process, but also enormous risks.

One of the most applauded and, in my view, accurate famous phrases attributed to the technologist par excellence of our era, Bill Gates, for the level of optimism but at the same

time of caution that is implicit in it, is the following: "*The first rule of any technology used in a business is that automation applied to an efficient operation will magnify efficiency. The second is that automation applied to an inefficient operation will magnify inefficiency*".

And I consider that such a statement, full of content and force, applicable, no doubt, to the use of AI in both the public and private sectors, could be the perfect summary of what I have tried to expose throughout this doctoral thesis. Let me explain.

As can be deduced from the content of this paper, AI is, of course, an unprecedented technological tool that has more than enough capacity to magnify the efficiency of human tasks. However, this has an important nuance. As I have been explaining, this technology works through powerful systems that carry out the analysis of huge amounts of data to produce certain results. Thus, its success or failure depends, on the one hand, on the power and technical characteristics of such systems and, on the other hand, fundamentally on the quality of the data analysed.

Accordingly, for example, if in an organization, before the introduction of AI, human, non-automated decisions were being made on the basis of poor quality data (i.e., illegally obtained, unrepresentative, erroneous, etc.) and therefore incorrect and inefficient (given the incompetent and unhelpful use of resources), then the incorporation of such powerful technology as AI and the automation of tasks without any change in the quality of such data will only multiply and magnify the already historical inefficiency of the organization and its operations, which could have highly detrimental results for citizens, as it would perpetuate patterns that should certainly tend to disappear.

However, this apparently simple reflection, harbours a broad technical and legal debate that, especially in relation to the use of AI in the field of criminal investigation (and, specifically, in the Spanish investigative process), is the one that has given basis to the present doctoral thesis.

Already at the beginning of this work I tried to warn of the great difficulty of establishing a universal and permanent definition of what AI is without the risk of falling into obsolescence, due to the rapid and constant evolution to which such technology has been subjected.

However, it is true that throughout the same, the foundations of the aforementioned concept have been laid and the elements common to all those techniques that make up the aforementioned notion have been outlined.

And this, undoubtedly, is one of the issues that have been clearly evident after analyzing the different applications that AI has or may have in the field of criminal investigation, because no matter how much terminological and technical fuss underlies, the reality is that there is a common element in all of them that serves as a foundation and common thread, mainly for those who are laymen in technology, namely: the detection and recognition of patterns among huge amounts of data as a basis for automatic decision making.

Thus, the raison d'être of AI is (or should be, in my opinion) none other than problem solving by using as much information as possible (to which we should add: "quality"), in order to make the best and fastest decisions for the benefit of human beings.

However, as this paper has shown, unfortunately, this is not always the case.

The benefits that AI systems can bring to our societies in general are more than clear, namely, and mainly: the liberation of humans from automatable tasks and, therefore, the increase of their time and their ability to carry out activities that require essentially human qualities (worth the redundancy), namely creativity or empathy; and the increase (to sometimes unimaginable levels) of efficiency -mainly in terms of accuracy, time and costs- in decision making.

Specifically, in the field of criminal investigation, the potential benefits are also clear, namely, among others: increased efficiency in crime prevention; assistance to the authorities -judicial police, prosecutors and judges- in making decisions based on the prediction of future events (especially in the adoption of precautionary measures); increased efficiency of material and personal resources (which translates, essentially, into an increase of the quality and speed of the resolution of cases and a reduction of costs); and an unprecedented increase of the success rates of the investigation of cases.

However, as has already been pointed out on several occasions throughout this doctoral thesis, the potential risks involved in the use of AI systems cannot be lost sight of at any time, given

that they have the capacity to far outweigh the benefits listed and have a multiplying effect on the historical weaknesses present in both police and judicial procedures, as Bill Gates warns in the famous phrase that begins this Section.

Thus, in general, I believe that humans should be alert to the possible presence of, among others, the following risks: the massive infringement of fundamental rights and individual liberties of citizens as a consequence of the indiscriminate use of AI systems; the loss of human control of AI systems and thus the potential breakdown of the principle of using them for the benefit of humanity; widening the gap between majorities and minorities and rich and developing countries; the lack of transparency and explainability of AI systems and, therefore, the loss of citizens' rights and trust; technical difficulties and the existence of security breaches leading to irreparable consequences for human beings; problems related to liability for the eventual misuse or malfunctioning of AI systems and the eventual lack of protection for the citizens affected; or the loss of jobs and the consequent impoverishment of society. And, in relation to the latter, it is very graphic the example set already in 2015 by the scientist Stephen Hawking, who warned: "*Robots could make us rich and free, but it is more likely that we will end up poor and unemployed*".[811]

Specifically, in the field of criminal investigation, we must also be especially attentive to a possible transgression of fundamental rights as essential as the dignity of the person, freedom, equality and non-discrimination, the right to defense and the right to honor, and privacy and protection of personal data, among others; a potential use of AI systems to carry out the perpetration of crimes in a more sophisticated and unpunished way; a limitation of the right of access to justice depending on the group or country to which each individual belongs; the impossibility of knowing if a legal process has been carried out with all the guarantees, of knowing if there have been violations of rights, and of refuting the arguments that have led the police, prosecutors and/or judicial authorities to take certain decisions, due to lack of transparency; the helplessness of citizens affected by the misuse or malfunctioning of AI systems; the increase in unemployment among justice officials and police, prosecutors and judicial authorities; and, above all, the most worrying aspect: the dehumanization of justice.

[811] See Bolton, 2015.

And the fact is that such risks, which in themselves, in general, involve serious dangers for human beings, in the field of criminal investigation are multiplied and are absolutely intolerable and contrary to all constitutional and legal principles in force in our country, as I have already shown throughout this doctoral thesis.

SECOND.- The only way that AI can be used by the authorities in the field of criminal investigation without carrying out massive violations of rights is its regulation.

Thus, after a thorough analysis of these dangers, I have come to the conclusion that the only possible solution to use AI in the field of criminal investigation in a safe and secure way is to try to minimize such risks through regulation. A regulation that, on the one hand, must define and delimit very well in which cases AI systems can be used for criminal investigation, and in what terms; on the other hand, it must establish mechanisms to guarantee their quality both at the time of their commercialization or release and afterwards; and, finally, it must contain elements of control of compliance with the legal requirements at all times.

Already in the 19th century, the German jurist Frederick Charles de Savigny, a convinced Romanist, warned of the disadvantages and problems that codification could cause, given that social reality is highly changeable and modifying laws is neither an agile nor a light task, and that is why he understood that regulation could imply that the Law would always be one step (or more) behind reality. Such an argument, no doubt, in my opinion, is largely correct, and even more so in a field such as AI, which varies and evolves at great speed. However, I believe that the vision of another German jurist, Anton Friedrich Justus Thibaut, who saw the need to bring order to legislation, complete it and establish clear legal bases in order to put an end to what I believe to be the dangerous "legal particularism" and make way for legal certainty, is also ideal.

We have already seen throughout this paper that the incomplete and inadequate legislative maze that we currently face when talking about AI, especially in the EU (since there is no specific and clear regulation), leads us to try to guess which legal instrument is applicable in each case, often without success, which opens the door to the existence of legal loopholes and confusing situations that do nothing more than cause distrust and lack of protection for citizens.

Thus, in my view, an essential precondition for the use of AI tools in any case, and especially in the field of crime investigation, is to legislate, since this is the only way to generate legal certainty and confidence in citizens and, above all, to avoid massive violations of fundamental rights. And I understand that the most suitable, in this sense, and in order to act in balance between what both Savigny and Thibaut advocate, would be to establish a "minimum" regulation, which should be clear and specific but which would leave the door open to examine each specific case in a flexible way, precisely because of the constant and rapid evolution that AI is undergoing.

At present, however, in the EU and in Spain, unfortunately, this has not taken place, which has left the door open to numerous violations of rights and freedoms that have been constantly denounced by pro-human rights organizations, which have “taken the authorities to task”, most of the time with good reason.

Faced with this problem, however, and given the barbarities that have been committed in recent years (and continue to be committed) through the use of AI systems without having a clear and proper regulation, the authorities have finally become aware of the need to establish a clear legal framework applicable to the design, development, marketing, distribution and use of AI systems.

THIRD.- The necessary elaboration of a Universal Declaration of Principles on AI.

In relation to this, and regardless of the fact that each country (or group of countries, as in the case of the EU), in the exercise of its own sovereignty, proceeds to develop specific legislation on such issues (which should reflect the principles and values that the legislature deems appropriate at all times), as I have already shown throughout this work, I believe that it would be very necessary and beneficial for humanity to establish basic principles of AI, a kind of "Bill of Rights", similar to what was done on December 10, 1948 with the Universal Declaration of Human Rights, which would guarantee ethical and moral minimums that the systems that use such technology should comply with at the international level and would stand as a real shield of protection against this new, enveloping and dangerous technology.

For this, however, the framework of an international organization such as the UN or any other public institution, pre-existing or created *ad hoc*, would be needed to bring together the largest possible number of signatory countries and, desirably, to make such a declaration of principles binding, establishing the corresponding sanctions for cases of non-compliance and establishing a highly powerful system of prevention and deterrence of infringement, since it must be taken into account that in case of violation of the agreed bases, it would be practically impossible to recover control and effectively purge responsibilities in an effective way.

And that, no doubt, in my view, is a mere question of willingness and compromise. Willingness to renounce to a little of one's own national sovereignty in the short term in exchange for preserving it as a species in the future; willingness to limit national productivity and wealth in exchange for maintaining freedom as a group of human beings; and commitment not to break the rules for the benefit of all, in order to guarantee survival. And so, in my view, if clear and binding basic principles are not established globally to guide everyone in the use of AI, the risk that such technology will end up beyond human control and denature us as a species is real. And we have been warned.

In relation to this, it is clear that not all countries have the same interests and the same facility to reach agreements with others (since some of them are even enemies), but in this case I think that the world leaders should make an extra effort and look "beyond their noses" to adopt commitments of international significance that would undoubtedly revert to their own benefit in the future, since the alternative in my point of view is extremely dangerous. And the worst thing is that it is enough for a country with certain power to do things badly for the rest (and itself) to be harmed, so the pact should, in my opinion, aspire to be signed by all existing countries, which, moreover, would imply a minimum guarantee of equality that could lead to overcome the current gap that stalks the world today (especially the African continent, which is at the bottom) in terms of AI.

Unfortunately, however, and despite the large number of initiatives and projects that are underway to regulate AI, the above still seems utopian, since it does not seem that, at least for the moment, we can have real expectations about the creation of a Universal Declaration of AI Principles that would be binding for almost all countries in the world, so we must be satisfied with mere declarations of principles with a guiding value that, hopefully, will at least

serve to inspire national legislative powers (or transnational ones, as is the case of the EU) when it comes to establishing legislation on AI.

In any case, it is clear that we are facing an unprecedented phenomenon and, fortunately, society's awareness of the threat posed by AI and the need to set limits to it is growing, so I am reluctant to rule out the above option.

In relation to this, as I have already stated throughout this work, I understand that the basic principles that should be contemplated in the aforementioned proposal for a Universal Declaration of AI Principles would be the following:

- Principle of respect for the dignity of the human being, with guaranteed supervision and control, and priority to social and environmental well-being;
- Principle of respect for the freedom and privacy of the human being, with a guarantee of individual management of personal data, transparency and explainability of the systems;
- Principle of equity, equality, non-discrimination of human beings and inclusion;
- Principle of robustness, technical soundness and safety; and
- Principle of responsibility.

FOURTH.- The imperative need to develop national (or supranational) laws to ensure responsible and safeguards the use of AI in the field of criminal investigation.

Having said this, despite not having at present, unfortunately, a common basic legal framework, I believe that it is urgent that the different countries proceed to regulate AI within the scope of their own sovereignty or that of the organizations of which they are part (such as, for example, the EU). I believe that, until now, the lack of knowledge and fear of this technology, together with the hope that common principles of action would be established at the international level, have been the main cause of the passivity of the States with regard to the establishment of clear and specific national legislation on this technology. However, now that the use of AI is the order of the day and the absence of regulation is often causing massive violations of citizens' rights and freedoms, as mentioned above, the need for legislation is urgent. Indeed, we are already late.

I am particularly critical of the EU, which, although, as I have explained in this doctoral thesis, has recently been taking firm steps towards the regulation of AI, the truth is that it has been "asleep" during a key period and, as a result, has been overtaken by countries such as the USA, China or Israel in relation to the use of such technology, which seems to me to be particularly serious, since I understand that the EU, the largest single market in the world, had the responsibility to develop a regulation that would be in line with its principles and values that would guarantee an AI use respectful of fundamental rights.

The EU's power of influence is enormous and undoubtedly goes beyond its borders, given that the countries producing such technology (which are usually those that are least respectful of human rights), should they wish to market their products within the Union, would have to adapt to the standards set by the Union, which would deter them from the outset, to a large extent, from developing systems that do not comply with the values and principles of the EU, which are highly protective of the rights and freedoms of citizens.

However, my disappointment goes even further, since in the EU Member States themselves, have been using AI tools which, as I have already warned before, have been repeatedly denounced by human rights organizations for their potential infringements of the rights and freedoms provided for in the EU Charter of Fundamental Rights, which is absolutely unacceptable.

And I understand that legislating at an EU level is neither easy nor quick, but certainly its heavy machinery cannot go against the rights and freedoms of its citizens, since it is absolute nonsense. And I say this when, to this day, the final text of the European Regulation on AI has not yet been approved (and it is not expected to be published until 2022), which continues to cause tremendous legal uncertainty and lack of protection for citizens, since although there is a White Paper on AI, it is not enough, and the current legal muddle and disorder are taking their toll.

Spain, for its part, even though it has the sovereignty to do so, has not decided to regulate the use of AI, although it had the guide of basic principles established in the aforementioned European White Paper on AI. However, I assume that now the Government is already waiting for the publication of the Regulation on AI, which will be directly applicable at the national

level. And the fact is that our country, unfortunately, did not even present a National AI Strategy until 2021, despite the fact that countries around the world already started to present theirs in 2017 (the first was Canada), which is absolutely unjustified and very disappointing, especially because in November 2017 the Spanish Government published the constitution of a Group of Wise Men on AI and Big Data[812] for the publication of a Spanish White Paper on AI, which, however, never came out.

Despite the above, however, as I have already said, I do not lose hope. And, of course, one of the reasons why I decided to do this doctoral thesis was precisely to study in depth the possibilities offered by AI in criminal investigation in order to be useful, eventually, to the EU or to our country, if it is finally decided to establish a clear and specific regulation in this regard, since I know that the field of criminal investigation through AI, unfortunately, is not very well explored because it is not very profitable. It is well known that where a greater investment and a higher level of technological research is carried out is in the private sector, not in the Administration, which unfortunately does not usually allocate large resources to such matters (and, if it does, then it usually does not achieve efficient and successful results).

As it has been explained throughout this paper, the utilities of AI tools for investigating crimes (both predictive and risk assessment tools, as well as criminal investigation tools themselves, and processing tools) entail, of course, an unprecedented revolution in the field of criminal investigation.

The use of systems capable of diving into millions of data per second and assisting the judge in making decisions in a precautionary manner; the use of systems that can identify or verify the identity of a person through the analysis of their biometric data; the use of systems capable of interacting with victims or witnesses and execute spoken commands, as well as filtering and organizing the information contained in millions of documents or detecting false allegations; the use of tools capable of "cyber patrolling", or of simultaneously translating and transcribing judicial statements; the use of tools with the potential to analyze images, read license plates in real time or detect false documents; or the use of tools capable of discovering

[812] Government of Spain, 2017.

frauds and swindles and detecting the production of gunshots is, without a doubt, a phenomenon that can increase the efficiency of criminal investigation to unsuspected levels.

However, as it has also been said, such benefits must be balanced against the potential risks that such systems may cause, and the real chances that the results of such systems will have a real influence on police, prosecutorial and judicial decision-making.

And the solution to this issue, however, in my view, as I have repeatedly announced, is very clear: regulation (without falling into excess) and, above all, the creation of mechanisms to ensure strict compliance.

In my opinion, then, it should be the legislature that, at each moment, decides how it wants to carry out the aforementioned balance between the possible benefits and potential risks involved in the use of AI tools for crime investigation (subject to and respecting, where appropriate, the common basic principles that could be approved at the international level), especially taking into account its political positioning, particularly with respect to the freedom-security equation.

To do so, however, I believe that the legislature (both European and national) must be advised in a cross-cutting manner by the most prestigious and brilliant experts in the field of AI, personal data protection, cybersecurity and criminal law, since legislating on a matter as complex and dangerous as the use of such technology in the field of criminal investigation, with the amount of fundamental rights at stake, must be carried out in the most protective and cautious way.

And it is well known that not everything that is technically possible has to be legally viable, which is why, in my opinion, multidisciplinary collaboration is absolutely necessary, with the capacity to assist the legislative power in the most technical, objective and correct way possible, in its task of legislating without prejudice to the fact that later, depending on its tendency at any given moment, it may choose to take the legislative decisions that it considers most convenient, more or less guaranteeing, or more or less flexible.

In any case, what I do consider essential is that the laws regulating the use of AI for criminal investigation be specific, complete and, above all, clear. And I emphasize this last requirement because such legislation should serve as a guide not only for the legal operators who apply it, but also for the designers and developers of systems that, of course, must create them so that they can comply with legal requirements. And that is what, in conversations with various engineers held to carry out this work, has been most commonly and vehemently conveyed to me, since according to what they tell me, it is very difficult for them to work without a clear legal guide for action and, as a result, sometimes they program systems that are then identified as illegal and are demonized (and, no wonder, because of the indignant legal vacuum and the legislative maze that exists today).

FIFTH.- The current European and Spanish steps towards the regulation of the use of AI in the field of criminal investigation are good news but they are late and sometimes insufficient.

Despite the fact that in the EU and in Spain, as I have already warned previously, we unfortunately do not currently have a clear and specific regulation on AI, it is appropriate to keep our attention, on the one hand, on what, as has already been said, will become the text of the future Regulation on AI, which is not yet definitive, as for the time being there is only the Proposal for a Regulation of the European Parliament and of the Council laying down Harmonized Rules in the field of Artificial Intelligence (Artificial Intelligence Act) and amending certain legislative acts of the Union; and, on the other hand, in the new text of the Spanish Criminal Procedure Act which is in the making, although for the moment there is no news that it will include a specific regulation on the use of AI systems in the field of criminal investigation, which on the one hand is bad news but on the other hand is prudent, given that we are awaiting the publication of the aforementioned European Regulation which will undoubtedly have implications in this area.

That said, as I have already announced in previous pages, I understand that the regulation of AI, specifically in the field of criminal investigation, which is the one that concerns us, in order to be useful, must, on the one hand, define and limit very well in which cases and under what circumstances AI systems can be used; on the other hand, establish mechanisms to ensure the quality of them both at the time of their commercialization or release and

afterwards; and finally, provide mechanisms to ensure compliance with the legal requirements in force.

Regarding the first of the aspects to be regulated (namely, the cases and terms in which AI systems may be used in the field of criminal investigation), it is interesting to note that the aforementioned Proposal for a Regulation on AI, on the one hand, establishes which systems are prohibited and which can be used in each case; and, on the other hand, determines under what circumstances, if any, their use must be carried out.

Thus, for example, as already explained in Section 3.2.2.2. (A.3), Article 5(1)(d) provides for the prohibition of authorities from using real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes, unless their use is strictly necessary to:

-the targeted search for specific potential victims of a crime, including missing children;
-the prevention of a specific, significant and imminent threat to the life or physical safety of natural persons or of a terrorist attack; or
-the detection, tracing, identification or prosecution of the person who has committed or is suspected of having committed one of the offences referred to in Article 2(2) of Council Framework Decision 2002/584/JHA, provided that the law in force in the Member State concerned imposes a custodial sentence or detention order for a maximum period of at least three years.

As in this case, the aforementioned Proposal establishes a series of limitations on the use of AI tools in the field of criminal investigation that are extremely useful, both for the creators and designers of systems, as well as for legal operators and authorities, who, if the proposed Regulation on AI is finally approved, will have a clear guide for action.

And, with respect to all those tools that are not prohibited, a series of legitimizing circumstances for their use and a series of mandatory requirements are foreseen, which is also extremely useful.

In relation to this, and following on from the previous case, it is provided, for example, that the use of real-time remote biometric identification systems in public access areas, in those cases in which they are legally permitted because they are strictly necessary to achieve one of the aforementioned purposes, must be authorised by a judge or by an independent administrative authority of the Member State where it is to be carried out. It also provides that such authorisation must be based on the legal provisions, if any, of the national legislation of each Member State, under the terms provided for in the Regulation itself.

And the fact is that the use of such systems, in particular, involves the risk that our daily lives become a kind of "Big Brother" that involves the absolute loss of, mainly, our rights to freedom and privacy, so it should certainly be limited, contrary to what is thought in some countries (China, for instance). Regarding this, we have to wonder: what would we think if by law it was allowed that a policeman proceeded to follow us every time we left home and witness all our movements? Of course we would, and that, surely, is exactly the same as having cameras placed on the street tracking our every move through remote biometric identification in real time. And that is why the EU is rightly seeking to restrict their use.

All this, however, must be put in relation to the current provisions of the legislation on the protection of personal data, which, although it does not directly and specifically regulate the use of AI systems, it does have implications in this respect, given that they carry out, on most occasions, the processing of special categories of data of this kind. And, specifically, it is appropriate to focus on the provisions of the Organic Law 7/21, of May 26th, on the protection of personal data processed for the purposes of prevention, detection, investigation and prosecution of criminal offences and the execution of criminal penalties, especially in its articles 13 and 14, as it has been shown in this doctoral thesis.

Therefore it is the case that AI systems, as already mentioned, are essentially fed on data, having been cataloged as “the oil of the XXI century”, by the enormous value they have achieved, since they are used to train such tools in order to improve their quality.

However, “the devil is in the data”, since, as it is already known, not just any type of data is good enough to achieve such an end, since the data that feed an AI system must be of good quality, because otherwise, the only thing that is achieved is, on the one hand, to perpetuate

historical patterns, in many cases illegal, and on the other hand, to create new unlawful situations that violate the rights and freedoms of citizens.

And, of course, as we have been seeing, the quality of AI systems depends on their technical characteristics but, above all, on the data they use, since they are the key element of such technology.

Thus, if an AI tool uses unrepresentative data, its levels of accuracy would tend to decrease and the risks of violation of the right to equality and non-discrimination would tend to increase; and if an AI tool processes data obtained illegally or carries out processing that does not comply with the legally established requirements, the risk of violation of the right to the protection of personal data, among others, would tend to increase, so it is essential, in my view, to pay special attention to the type of data used and, above all, to the treatment given to them.

And, of course, the above makes it absolutely necessary to be clear, *ex lege*, in what cases and under which circumstances the processing of personal data for the use of AI tools in the investigation of cases can be carried out, given that an error in this regard could have dire consequences. And, among other issues, what would happen if an AI tool used by a judge, when challenged by the defense, is declared to have been used illegally due to improper processing of personal data, for example? Of course, this would lead to the nullity of all decisions taken based on the results obtained (with the risk of causing irreparable damage), but, undoubtedly, it would also be very difficult for the judge to abstract from them, as it would happen, for example, with the content of an illegal evidence that cannot be officially assessed but is known unofficially. Without doubt, a real nonsense.

Regarding the second of the aspects to be regulated (namely, the mechanisms to guarantee the quality of AI systems for criminal investigation, both at the time of their commercialization or putting into circulation, as well as afterwards), it is interesting to point out that the already analyzed Proposal for a Regulation on AI establishes a series of procedures, requirements and controls that vary depending, especially, on whether or not they are classified as high risk AI systems according to the provisions of Article 6.

It is clear that all possible precautions must be taken to ensure that AI systems are used for the benefit of investigation without violating any rights and, above all, without generating anti-legal situations that may be irreversible.

In relation to this, as it has been said, the aforementioned proposal lays down procedures for assessing the conformity of AI systems classified as high risk[813] with the legislation in force, which I welcome. However, in my view, the proposed model is not the most appropriate and suitable for guaranteeing the existence of a real filter and the existence of a real institutional verification of the quality of the aforementioned tools. Let me explain

I understand that the conformity assessment system proposed in the aforementioned text is perhaps the most economically viable and the one that best fits in with the current resources and the current configuration of the European institutions, but, as I have been warning, I believe that AI is a matter that involves serious dangers and that, therefore, it must be treated with the greatest caution, even if it involves a greater investment, especially when the regulations themselves reference to systems classified as high-risk.

My proposal in this regard, which has already been announced on several occasions throughout this doctoral thesis, is none other than the creation, at the EU level, of an Agency on AI that has the competence to examine, filter and verify the quality of the systems and their conformity with current legislation, prior to their introduction on the market or their circulation and afterwards. In my opinion, such an agency could have a central headquarters and, in turn, delegations in the various Member States, in order to be able to decentralize these operations, to be carried out more swiftly and more easily, always under common and unique processes, ultimately controlled from the main headquarters.

This model, similar to the one followed by the European Medicines Agency, is in my opinion the most optimal for carrying out the tasks of analysis and verification of the quality and conformity of the systems, since, on the one hand, it would be a public institution that would be behind the process, which is essential in order to generate trust in citizens, since the only

[813] See pages 132 and 133.

interest is (or, at least, should be) the officially announced one; and, in addition, there is a very low risk of leakage of the millionaire patents that are usually behind such tools.

However, as it has been explained in Section 3.1., the Proposal for a Regulation on IA envisages procedures for verifying the quality of the systems that are different from the one I am proposing and, only exceptionally does it provide for the active intervention of the authorities, which, as I have said, does not seem to me to be the most appropriate thing to do.

And finally, with regard to the last of the aspects to be regulated (namely, the mechanisms to ensure compliance with the legal requirements in force), it should be pointed out that the analyzed Proposal for a Regulation on IA, as explained in Section 3.1., establishes in its Article 71 a series of sanctions for cases of non-compliance and also delegates to the Member States the creation of a system of sanctions for cases of violation of the provisions of the future Regulation, which I applaud and consider fundamental for the success of the legislation on such technology.

In addition to the above, the aforementioned Proposal includes some governance measures such as the creation of a European IA Committee to provide advice and assistance to the Commission in its task of coordinating and ensuring compliance with the provisions of the aforementioned body of law; and it also provides for each Member State to establish or designate competent national authorities to ensure the implementation and enforcement of the future regulation, which I consider essential, since I believe it is very important to ensure compliance with the legislation on AI, given the number of fundamental rights that could otherwise be violated and, above all, given that the EU's credibility in terms of guarantees of AI systems is at stake. And, in connection with this, in order to increase protection, moreover, quite rightly, in my view, a series of penalties are provided for in cases of non-compliance with the provisions of the future regulation.

In any case, and in order to establish prior controls and avoid infringements of rights, I consider it essential to create controlled testing areas for AI (the so-called “sandboxes”), which is quite rightly already provided for in the aforementioned Proposal. As stated in Article 53.1, these “sandboxes” provide a controlled environment that facilitates "*the development, testing and validation of innovative AI systems for a limited period before their introduction*

on the market or their putting into service, under a specific plan", which is highly beneficial, especially if carried out under the direct supervision and guidance of the competent authorities, as provided for in the said text.

SIXTH.- In any case, the use of AI by the authorities in the criminal investigation process must be carried out in a responsible and cautious manner, with special attention to the principle of transparency.

I must point out that, even if the use of AI systems for criminal investigation is legally limited and subject to quality controls, I understand that, given the enormous risks that they pose *per se*, the authorities using them must do so with the greatest possible diligence and caution, just as doctors do when prescribing medicines or pharmacists do when dispensing them, even though they are authorized for specific purposes by the European Medicines Agency.

Thus, in particular, I understand that the police, prosecutors and judges who would be entitled to apply AI systems for criminal investigation purposes, must carry out a very scrupulous and rigorous use of them, with full respect and guarantee in any case and under any circumstances to the right to defense. To this purpose, the people affected must have constant and absolute knowledge of the content of the tools applied, in order to be able to understand it and, if necessary, question and refute it, in application of the principle of contradiction, both during the investigation phase and in the plenary session. And this undoubtedly requires that the AI systems used in the field of criminal investigation comply with the strictest requirements of transparency, since this is the only way to determine what is behind the results that serve to make decisions that affect the people involved.

However, in relation to the latter, regarding the great controversy that, in my view, has rightly been generated around the need for transparency of AI systems, especially in cases of use in the field of criminal investigation, I would like to reiterate a reflection already made in the body of this doctoral thesis that seems important to me.

Although, unfortunately, nowadays, most of the time decision-making with algorithms becomes synonymous with inexplicable decision-making (which is often defined as a “black box”), and this is unacceptable, we must not forget that police officers, prosecutors and judges,

as humans, also have their own black boxes (namely, their brains) which, in fact, is much worse than those of AI systems, since they become inscrutable in any case.

In this regard, I understand that, although there are clear differences between the potential consequences of decision-making by a human and by a machine -especially in terms of the scope that can be achieved, which is multiplied exponentially if automated-, the truth is that both have common elements that should undoubtedly be taken into account and assessed, mainly in order to establish improvements, since that is what evolution is all about.

Thus, of course, not just because humans have their own *black box*, AI systems should be allowed to lack transparency (especially in the field of Public Administration), since, among other things, the authorities who make decisions in the field of criminal investigation always have the legal obligation to justify them legally (at least formally, without prejudice to the real motivations that have led them to resolve a matter in a particular way, which are almost impossible to discover), which is not the case with machines.

Accordingly, I believe that AI, in this sense, despite the current (and well-founded) doubts in this regard, may represent an unprecedented advance in the field of transparency in justice, given that, if a real filter that only allows the use of quality systems can be set, with guaranteed transparency, explainability and traceability, such a fundamental requirement will be taken to the highest level, with a clear improvement on current standards.

And this is especially important at a time when the idea of the so-called "robot judge", whose main purpose is to automate the resolution of *online* litigation through AI, is booming.

By way of example, among others, on the one hand, in 2017 the first “Internet Court” was opened in Hangzhou (China) to resolve disputes related to commercial issues (which today has a 99% acceptance rate of judgments and its average duration of holding trials and resolving cases is twenty-eight minutes).[814] Since then, such courts have also been set up in Beijing and Guangzhou, and between them, according to the report issued in December 2019 by the Chinese Supreme Court, they have resolved more than 100,000 cases since they were

[814] See Hangzhou Internet Court, n.d..

put into operation[815]. In addition, in March 2019, in twelve provinces of China, a "Mobile Court" was launched through the WeChat platform[816], and according to the aforementioned Chinese High Court in the aforementioned report, since its launch, it has resolved three million cases.[817]

On the other hand, in Estonia, already in August 2019 the government entrusted its Chief Data Officer, Ott Velsberg, with the coordination and supervision of the creation of an AI system that would allow the online resolution of minor disputes, of less than seven thousand euros, which today is already in pilot tests to become operational.[818]

SEVENTH.- Of all the risks inherent in the use of AI in the field of criminal investigation, there is one that is particularly dangerous: the dehumanization of justice.

In particular, however, and after referring to the potential risks of AI systems and their possible solution (namely regulation) I must say that of all the dangers outlined, there is one that, above the rest, particularly concerns me: that of the dehumanization of justice.

And the meaning of justice that best fits what I understand as such is the one that defines it as: "*Moral principle that leads to give to each one what corresponds or belongs to him*."[819]. However, as explained in Section 2.4.1., it is well known that morality is an abstract and indeterminate concept, whose content varies according to the beliefs and values of each person.

And what is so special about such an assertion? Well, precisely the singularity of the conception of morality and, therefore, of justice, that each of the individuals who inhabit the world have, and to my understanding, two identical ones do not exist. And the fact is that they may be similar or apparently the same, but never identical. And it is evident that police officers, prosecutors and judges are human beings, and as such, they also have their own concepts of justice.

[815] Chinese Supreme Court, 2019, p. 64.
[816] See Wang, 2019.
[817] Chinese Supreme Court, 2019.
[818] See The Technolawgist, 2019.
[819] Dictionary of the RAE.

It is clear that, nowadays, Spain is a state governed by the rule of law in which the law prevails over any kind of opinion or personal assessment when it comes to dispensing justice. However, the truth is that on many occasions the law gives the authorities the possibility to carry out interpretations, assessments of circumstances or exceptions based on their own experience or the particularities of the specific case. Yes, there is room for flexibility in justice. And this will be greater or lesser depending on the case, depending on the provisions of the applicable law and case law and, why not, depending on the concept of justice that has the competent authority that applies it at any given time. It would be absurd to deny that there are police, prosecutors and judges who are more and less rigid, more and less experienced, more and less empathetic, and more and less willing, since this is a reality.

Thus, from my experience as an investigation judge I can say that I know for a fact that some of the cases that have come before me would have been resolved in hundreds of different ways by my colleagues. And all of them would have acted with full respect for the law, but each one with an inclination towards their own concept of justice, always aiming to be as fair and objective as possible. And sometimes the law gives the power to do so. Thus, the more human side of justice undoubtedly exists.

How many times do police officers and prosecutors try to interact with the parties involved in a case and/or their lawyers so that they try to solve their problems, if possible, avoiding criminal proceedings to stop further damage? I, of course, in certain cases, for example, I refer the parties to mediation and/or psychological treatment if I see that this could be beneficial for the interested parties, and it often has an effect. How many times is it given a look of affection and understanding to a victim or offer words of encouragement? How many times is it tried to provide a medical remedy to people who are involved in real criminal loops because of their mental health problems or addiction to narcotic substances? How many times people are warned that if they continue like this they will end up in prison and try to motivate them to cut with their previous life and start a new stage? How many times the phone is picked up to talk directly to Social Services in cases of risk and try to solve issues that require human-to-human conversations? All of this happens, because we are people, and yet it is absolutely incompatible with automated decision making by a machine, based solely on algorithms.

On the one hand, it must be said that the emotional damage that interacting with a machine can cause to a person who is in a situation of maximum vulnerability (as occurs in practically 100% of cases in the field of criminal investigation) can be enormous, given that it generates immense impotence, since it already prevents a human with a certain empathy from assessing nuances that, for now, only a person can grasp (and life, precisely, is about nuances); and, on the other hand, it must be made clear that full respect for the right to the dignity of the person (which is the pure essence of the human being and what defines us as a species), could be undermined in the event of eliminating human intervention, which, in my view, is absolutely unacceptable. Not only because it denaturalizes us as a species, but also because I truly believe that it would not, in general terms, bring any better results than the current system.

Fortunately, however, in most countries of the world, as has already been shown in this work, the idea that AI should be focused on the interest of the human being, should not fully replace personal relationships and should always be controlled and supervised by a person, is strongly rooted. And in this sense, fortunately, all the projects, initiatives and regulations referred to in this doctoral thesis are moving in this direction, since humanity knows that there is a lot at stake behind the use of AI. And based on this premise, in my point of view, the use of AI in the field of criminal investigation should be focused on establishing itself as a suitable complement to the authorities, without replacing them.

Thus, by way of example, Article 14 of Organic Law 7/21, of May 26th, on the protection of personal data processed for the purposes of the prevention, detection, investigation and prosecution of criminal offences and the execution of criminal penalties, already mentioned above, generally prohibits decisions based solely on automated processing, including profiling, which produce negative legal effects or significantly affect him or her, unless expressly authorized by a regulation with the status of a law or by European Union law. And to this is added, moreover, that the rule enabling the processing "*must establish appropriate measures to safeguard the rights and freedoms of the data subject, including the right to obtain human intervention in the process of reviewing the decision taken*." This is undoubtedly in line with respect for the aforementioned values.

Notwithstanding this, the COVID-19 pandemic has shown the world more clearly than ever that investment and commitment to technology and, in particular, to AI, can be the perfect

ally for human beings if it is used correctly. And as we have been saying, the advantages of using AI systems in most areas of life are enormous. What cannot be lost sight of, however, although it is sometimes tempting, is that their design, development and use must respect, defend and promote the values and the rights and freedoms of the citizens of the societies in which they operate, while ensuring that this does not, however, hinder prosperity and competitiveness (a balance that is undoubtedly difficult to achieve).

EIGHTH.- Final conclusion: AI can be very useful to improve the efficiency of the Spanish investigative process only if its use is legally foreseen. Brief reflection on the evidentiary value of the proceedings carried out with AI tools.

In relation to all of the above, then, and by way of a final conclusion, after having carried out the research contained in the body of this doctoral thesis, I am in a position to answer the following question, configured as the hypothesis that has given rise to this work: can AI be useful for investigating crimes in the Spanish investigative process? The answer, clearly, is "yes", but with nuances. These nuances, however, as I have been explaining, mainly translate into potential risks that can be mitigated through the creation of basic and common principles of action, at the international level, and the development of clear, specific and complete national laws (or transnational laws, as in the case of the EU).

And in view of this, inevitably, the question arises as to the legal value, especially evidentiary, that could be granted today in Spain to the AI tools used by the authorities in the criminal investigation process, which, however, could give rise to the development of a complete independent doctoral thesis and exceeds the scope of the present one. Nevertheless, I consider it important to make a brief reflection on the subject.

As has been repeatedly noted throughout this paper, in our country we currently lack a special regulation on the use of AI by the authorities in the field of criminal investigation. However, we do have legislation on the protection of personal data, which, as has also been explained, contains some references that are applicable to such systems, and we also have a Spanish Constitution and a Charter of Fundamental Rights of the EU that are not only clear but have also been extensively interpreted by case law. And it is these legal instruments which, of course, I believe should operate today as a guide to, on the one hand, the use and application

of AI systems by the authorities in the criminal investigation process and, on the other hand, to determine their value. Let me explain.

In the absence of laws that specifically provide guidelines for legal operators and authorities (including prosecutors and judicial authorities) to use the aforementioned tools in criminal investigation, it is appropriate to resort to our basic principles and possible laws relating to other matters that may be applicable to the specific case. Since, unfortunately, we do not yet have a public system for certifying the quality of AI systems that allows them to be used with all the guarantees, we must take all the necessary precautions to avoid infringements of fundamental rights.

On the one hand, for example, if in a criminal proceeding the police were to present the judge with a report showing the use of a "cyberpatrol" AI tool that had detected the presence on the Internet of several suspected terrorists ready to carry out an attack, this should not automatically be taken as true and correct by the investigating judge, since, as is well known, there are many latent risks behind this kind of software. However, it would also be unwise to dismiss the information provided by the system out of hand, given that they have also been shown to work on occasion.

Accordingly, I consider that in such cases what should be done is a weighing of rights in order to determine, in each specific case, the value that should be given to the results presented by the AI tool used.

In order to do so, mainly, I understand that the origin and content of the AI system used should be taken into account, with the obligation to find out which companies are behind its design and marketing, what guarantees it offers, what filters the Administration has passed in order to be used, what levels of precision it manages and, above all, what degree of transparency and explainability it has.

And that, unfortunately, at present it is very difficult for an AI system (most likely created and distributed by a foreign *software* company and acquired by Spanish authorities) to have satisfactory answers to all these questions.

Thus, in view of the above, I consider that the investigating judge, in such a case, should require, without exception, that the AI tool used by the police be transparent and explainable, in order to be able to analyse, understand and contrast its content. In addition, however, I believe that the judge should require *additional* evidence in order to take into account the information submitted by the police, who should collect, based on the data obtained by the AI system, the maximum possible data to incriminate the suspects in order to provide the judge with as much information as possible so that he can give the results presented, at least, the value of *notitia criminis*.

And, of course, any judicial decision based solely on the result of an opaque *software* would be null and void (even if it were later proven to be right), since this would be contrary, in any case, to the right to a due process and the right to the presumption of innocence, without prejudice to the subsequent violations of rights that its use could entail (among others, right to defence, right to privacy, right to freedom, etc.). And any type of judicial decision based solely on the result of a transparent AI system whose quality is not publicly certified, given its possible lack of accuracy, the possible poor quality of its data and the consequent existence of biases, the potential presence of security breaches and, where appropriate, the difficulty of accountability, would also be potentially null and void due to the rights that it could violate.

And, likewise, in the event that a facial recognition system were to be used by the investigating judge, for example, to identify the perpetrator of a crime under investigation, I understand that the same standards would apply.

And it is that the *conditio sine qua non* so that, at least, the result thrown by an AI tool could count with value of *notitia criminis* and, even, diligence of investigation or evidence, I consider that it would be its transparency and explainability, since only with such characteristics the parties involved, the Public Prosecutor's Office and the judge could have real knowledge of the content of the algorithms to, if necessary, submit them to contradiction (in order to detect the possible existence of failures, errors, illegal data, security breaches and other issues that could lead to the nullity or loss of value.

However, all of the above will most likely change when the European Regulation on AI is finally adopted, which I hope will establish clear guidelines not only for the use of AI systems

by authorities in criminal proceedings but also for determining their legal value and, above all, finally establish a public system of quality certification that will allow authorities to use AI tools without an insurmountable fear of causing harm to rights. All of which certainly motivates me to carry out postdoctoral research in the future.

I am sure that, once a clear legal framework is established that delimits the use of AI in the field of criminal investigation and guarantees high quality standards, the authorities, legal operators and citizens will gradually increase their confidence in this new technology, and its use will become more widespread until it ends up becoming a common support element for police, prosecutors and judges.

In this regard, Richard Berk, the professor emeritus of criminology at the University of Pennsylvania (Philadelphia, USA) who designed the algorithm used by the Philadelphia Probation Department, has already predicted that controversies over the use of AI systems will fade as the algorithms become more widespread, and compared them to the autopilot systems now used in commercial aircraft, telling The New York Times: "*We've learned that the autopilot is reliable, more reliable than an individual human pilot. It's going to be the same thing here*."[820]

And, in my opinion, in relation to what the aforementioned professor says, the same thing that is happening today with AI, already happened in the past with each and every one of the new technologies that have been emerging, having been, in fact, described AI as the new electricity.[821] And so it happened, for example, with light, with airplanes, with cars, with the telephone, with the television, with computers and with the calculator, inventions that were seen almost as diabolical when they first appear and that today in a natural way, duly regulated, are part of our daily lives.

Thus, we must take our time to assimilate, understand and, above all, create the right mechanisms to ensure the use of AI in the safest and most reliable way possible, but without overdoing it.

[820] See Metz & Satariano, 2020.
[821] See Shana, 2017.

And, in this regard, it is appropriate to bring up the words of the wise writer Gabriel García Márquez, who said: “Since the appearance of visible life on Earth, it took three hundred and eighty million years for a butterfly to learn to fly, another one hundred and eighty million years to make a rose with no other commitment than to be beautiful, and four geological eras for human beings to be able to sing better than birds and die of love.”[822]

Without a doubt, a new exciting era is ahead of us.

[822] García, 1986.

5- BIBLIOGRAFÍA

5.1. PUBLICACIONES CIENTÍFICAS

Agustina, J. R., & Vargas, A. (2019). ¿Es necesaria una dogmática de los ciberdelitos? A propósito de la utilización de agentes encubiertos en la lucha contra la explotación sexual de menores en el ciberespacio. *Derecho y Persona.* Págs.609-644. Ideas Solución Editorial.

Barrio Andrés, M. (2018). *Derecho de los robots.* Wolters Kluwer.

Bathaee, Y. (2018). The Artificial Intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology, 31*(2). Págs.891, 901 y 905-906.

Berman, J. (2013). *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information.* Morgan Kaufmann.

Blackman, R. (15 de octubre de 2020). A Practical Guide to Building Ethical AI. *Harvard Business Review*.

Boden, M. (2016). *AI. Its Nature and Future.* OUP Oxford.

Boden, M. (2017). *Inteligencia Artificial.* Madrid. Turner.

Boix Palop, A. (2020). *Revista de Derecho Público: Teoría y Método, 1*. Págs. 223-270.

Brennan, T; Dieterich, W; Ehret, B. (Enero de 2009). Evaluating The Predictive Validity Of The Compas Risk And Needs Assessment System. *Criminal Justice and Behavior, 36* (1). Págs. 22-23.

Calderón Cuadrado, M. (31 de mayo de 2011). El derecho a un proceso con todas las garantías (aspectos controvertidos y jurisprudencia del Tribunal Constitucional). *Cuadernos De Derecho Público* (10). Pág.158.

Chalmers, D. (2010). The Singularity: A Philosophical Analysis. *Journal of Consciousness Studies* (17). Pág. 6.

Citron, D. (2007). Technological Due Process. *Washington University Law Review*. Págs.1.013-1.305.

Cohen, J. (2013). What is Privacy For? *Harvard Law Review* (1904). Págs. 1920-1921.

DeMichele, M., Baumgartner,, P., Barrick, K., Comfort, M., Scaggs, S., & Misra, S. (2019). What Do Criminal Justice Professionals Think About Risk Assessment at Pretrial?. *Federal Probation Journal, 83*.

Descartes, R. (1637). *El discurso del método.* Versión Kindle.

Dupuy, D., & Corvalán, J. (2021). *Cibercrimen III. Inteligencia Artificial. Automatización de algoritmos y predicciones en el Derecho Penal y procesal penal.* Editorial B de F.

Economou, N. (3 de octubre de 2017). A "principled" Artificial Intelligence could improve justice. *ABA Journal*.

Emerging Technology from the arXiv. (16 de noviembre de 2018). Machine Learning, meet quantum computing. *MIT Technology Review*.

Eterno, J., & Silverman, E. (2012). *The Crime Numbers Game: Management by Manipulation.* CRC Press.

Farley, B., & Clark, W. (1954). Simulation of Self-Organizing Systems by Digital Computer. *IEEE Transactions on Information Theory, 4* (4). Págs. 76-84.

Feldman, L., Adolphs, R., Marsella, S., Martinez, A. M., & Pollack, S. D. (17 de julio de 2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *SAGE Journals, 20*. Págs. 1-68.

Férez-Mangas, D., & Andrés-Pueyo, A. (2018). Eficacia predictiva en la valoración del riesgo del quebrantamiento de permisos penitenciarios. *LA LEY Penal, 134.* Pág.6.

Ferguson, A. (2017). *The Rise of Big Data Policing.* New York University Press

Fernández, C. (13 de febrero de 2020). Primera sentencia europea que declara ilegal un algoritmo de evaluación de características personales de los ciudadanos. *Diario La Ley*.

Galton, F. (1892). *Fingerprints.* Londres, Reino Unido: McMillan & Co.

García Moreno, J. (15 de junio de 2014). Reflexiones sobre el valor probatorio del reconocimiento en rueda. *Revista de Jurisprudencia* (2). Pág.8.

Gardner, H. (1983). *Frames of Mind: the theory of multiple intelligences.* Nueva York, EEUU: Basic Books.

González-Espejo, M. (19 de febrero de 2020). Sector público y algoritmos: Transparencia o un poco más de paciencia. *Diario La Ley*.

Hao, K. (21 de enero de 2019). AI is sending people to jail-and getting it wrong. *MIT Technology Review*.

Jastrow, R. (1985). *El Telar Mágico*. Salvat Editores.

Kaplan, J. (2017). *Inteligencia Artificial. Lo que todo el mundo debe saber.* Teell.

Kingston, J. (15 de marzo de 2018). Pautas legales para juzgar a una Inteligencia Artificial asesina. *MIT Technology Review*.

Knight, W. (17 de diciembre de 2015). Mil millones de dólares para evitar que la IA sea 'mala' con la humanidad. *MIT Technology Review*.

Knight, W. (25 de octubre de 2017). AlphaGo Zero ha derrotado a su hermano mayor en 100 a 0 sin ayuda humana. *MIT Technology Review*.

López de Mántaras Badia, R., & Meseguer González, P. (2017). ¿Qué sabemos de? Inteligencia Artificial. CSIC y Catarata.

Markoff, J. (2016). *Machines of loving grace: The Quest for common ground beyond humans and robots.* Ecco.

Melé, D., & Sánchez-Runde, C. (2013). Cultural Diversity and Universal Ethics in a Global World (2013). *Journal of Business Ethics*. Págs. 681-687.

Minsky, M. (Octubre de 1994). Will Robots Inherit the Earth? *Scientific American*.

Necati Pehlivan, C., Cervera Navas, L., Cuatrecasas Monforte, C., Keser Berber, L., Atabey, A., & otros. (Febrero de 2020). Legal Challenges of Artificial Intelligence (AI) . *Global Privacy Review, 1*. Pág. 6.

Newell, A., & Simon, H. (1961). *GPS: A program that simulates human thought.* Computers & thought. MIT Press. Págs. 279-293.

Nilsson, N. (2010). *The Quest for Artificial Intelligence: A History of Ideas and Achievements.* Cambridge University Press.

Ortega Klein, A. (2015). *La imparable marcha de los robots* . Madrid. Alianza Editorial.

Oswald, M., Grace, J., Urwin, S., & Barnes, G. (2018). Algorithmic risk assessment policing models: lessons from the Durham HART model and "Experimental" proportionality. *Information & Communications Technology Law, 17.* Págs. 223-250.

Parasuraman, R., & Manzey, D. (Junio de 2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *The Journal of the Human Factors and Ergonomics Society, 52.* Págs. 381-410.

Picó i Junoy, J. (2012). *Las garantías constitucionales del proceso II Parte. Análisis del art. 24 C.E. a la luz de la doctrina del Tribunal Constitucional. Derecho a no declarar contra sí mismo y a no confesar culpable.* J.M. Bosch Editor.

Richardson, R., Schultz, J., & Crawford, K. (2018). *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice.* New York University Law Review Online. Última visita el 21 de febrero de 2021. https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson_etal-FIN.pdf

Rifkin, J. (2011). *La tercera Revolución Industrial: cómo el poder lateral está transformando la energía, la economía y el mundo.* Paidós.

Rochester, N., Holland, J., Habit, L., & Duda, W. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IEEE Transactions on Information Theory, 3*. Págs. 80-93.

Rogel Vide, C., Lacruz Mantecón, M., Mozo Seoane, A., & Diaz Alabart, S. (2018). *Los robots y el Derecho. Jornadas*. *Colección Jurídica General*. Reus Editorial.

Rudin, C., & Radin, J. (22 de noviembre de 2019). Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition. *Harvard Data Science Review, 1.* Pág. 2.

Schaw, K. (2016). *La cuarta Revolución Industrial.* Penguin Random House.

Searle, J. (1980). *Minds, brains, and programs.* Cambrigde University Press.

Singer, P. (2009). *Wired for war. The robotics revolution and conflict in the 21st century.* Nueva York. The Penguin Press.

Smith, B., & Shum, H. (2018). *The Future Computed Artificial Intelligence and its role in society.* Washington. Microsoft Corporation Redmond.

Stevenson, M., & Slobogin, C. (2 de agosto de 2018). Algorithmic Risk Assessments and the Double-Edged Sword of Youth. *Washington University Law Review, 96*. Págs. 18-36.

Stewart Morrison, G., Fyder Sahito, F., Jardine, G., Djokic, D., Clavet, S., Berghs, S., & Goemans Dorny, C. (junio de 2016). Interpol survey of the use of speaker identification by law enforcement agencies. *Forensic Science International, 263*. Págs. 92-100.

Taddeo, M., & Floridi, L. (24 de agosto de 2018). How AI can be a force for good. *Science, 61*. Págs. 751-752.

Teigens, V., Skalfist, P., & Mikelsten, D. (2019). *Inteligencia Artificial: la cuarta revolución industrial.* Cambridge Stanford Books.

Uchida, C., & Swatt, M. (2013). Opeation LASER and the Effectiveness of Hotspot Patrol: A Panel Analysis. *SAGE Journals, 16.* Págs. 287-304.

Van der Sloot, B., Broeders, D., & Schrijvers, E. (2016). *Exploring the Boundaries of Big Data.* Amsterdam University Press.

Waldman, A. (29 de septiembre de 2019). Power, Process, and Automated Decision-Making. *Fordham Law Review, 88*. Pág.5.

Wallach, W., & Allen, C. (2010). *Moral machines: Teaching robots right from wrong.* Oxford University Press.

Wilkins, N. (2019). *Inteligencia Artificial: Una Guía Completa sobre la IA, el Aprendizaje Automático, el Internet de las Cosas, la Robótica, el Aprendizaje Profundo, el Análisis Predictivo y el Aprendizaje Reforzado.* Edición de Kindle.

5.2. PUBLICACIONES EN PRENSA

Abad Liñan, J. (6 de noviembre de 2015). La voz, prueba contra el crimen machista. *El País*.

Adams, T. (12 de junio de 2016). Artificial intelligence: "We're like children playing with a bomb". *The Guardian*.

Agerholm, H. (16 de noviembre de 2018). Metropolitan police 'gangs matrix' breached data protection laws, investigation reveals. *The Independent.*

Alandete, D. (27 de octubre de 2011). John McCarthy, el arranque de la Inteligencia Artificial. *El País*.

Álvarez, D. (9 de diciembre de 2018). Inteligencia Artificial al servicio de la técnica forense. *El Bierzo Digital.* Última visita el 2 de marzo de 2021. https://www.elbierzodigital.com/inteligencia-artificial-al-servicio-de-la-tecnica-forense/268294

Barbieri, A. (9 de febrero de 2019). EE.UU. crea un algoritmo que predice golpes de estado y crisis financieras. *La Vanguardia*.

Barbieri, A. (18 de marzo de 2019). ¿Puede predecirse un crimen antes de suceder con un algoritmo? *La Vanguardia*.

Bejerano, P. (2 de mayo de 2019). "Automa cavaliere": el robot que diseñó Leonardo Da Vinci. *El País*.

Benedito, I. (21 de febrero de 2020). EEUU avisa a España: no compartirá inteligencia si utiliza 5G de Huawei. *Expansión*.

Benner, K. (6 de febrero de 2020). China's dominance of 5G Networks puts U.S. economic future at stake, Barr warns. *The New York Times*.

Berbell, C. (29 de noviembre de 2020). Así será el futuro: Juicios y declaraciones transcritas por inteligencia artificial en pantalla y descargables en pdf o word. *Confilegal*. Última visita el 23 de septiembre de 2021. https://confilegal.com/20201129-asi-sera-el-futuro-juicios-y-declaraciones-transcritas-por-inteligencia-artificial-en-pantalla-y-descargables-en-pdf-o-word/

Bolton, D. (9 de octubre de 2015). Stephen Hawking says robots could make us all rich and free. *The Independent*.

Campbell, D. (7 de julio de 2013). Bajo la vigilancia de los Cinco Ojos. *El País*.

Castro, C. (30 de agosto de 2020). La justicia sigue en la picota. *La Vanguardia*.

Certes, N. (21 de mayo de 2018). Le code source de Parcoursup enfin publié. *Le Monde Informatique*. Última visita el 3 de agosto de 2020. https://www.lemondeinformatique.fr/actualites/lire-le-code-source-de-parcoursup-enfin-publie-71802.html

Chandrasekaran, R. (12 de mayo de 1997). Kasparov Proves No Match for Computer. *The Washington Post*.

Chokshi, N. (25 de febrero de 2020). Tesla Autopilot System Found Probably at Fault in 2018 Crash. *The New York Times*.

Conger, K., Fausset, R., & Kovaleski, S. (14 de mayo de 2019). San Francisco Bans Facial Recognition Technology. *The New York Times*.

Corona, S. (8 de abril de 2018). La robot Sophia: “Los humanos son las criaturas más creativas del planeta pero también las más destructivas”. *El País*.

CNN Business. Clearview AI's founder defends controversial facial recognition app. Última visita el 21 de mayo de 2021. https://edition.cnn.com/videos/business/2020/02/10/clearview-ai-facial-recognition-orig.cnn-business

Criado, M. (25 de octubre de 2018). ¿A quién mataría (como mal menor) un coche autónomo? *El País*.

Darrach, B. (20 de noviembre de 1970). Meet Shaky, the first electronic person: The fascinating and fearsome reality of a machine with a mind of its own. *Life*. Pág. 66.

Duhigg, C. (16 de febrero de 2012). How Companies Learn Your Secrets. *The New York Times*.

Fernandez, M. (2 de julio de 2020). Mercadona usa reconocimiento facial para detectar a quien no puede entrar en sus tiendas. *El Español*.

Fresneda, C. (10 de junio de 2014). Un ordenador logra superar por primera vez el test de Turing. *El Mundo*.

Glez, M. (25 de octubre de 2018). Nosotros los robots. *El País*.

Gorner, J. (21 de agosto de 2013). Chicago police use “heat list” as strategy to prevent violence. *Chicago Tribune*.

Harwell, D. (17 de julio de 2019). FBI, ICE find state driver’s license photos are a gold mine for facial-recognition searches. *The Washington Post*.

Hawkins, S. (2 de diciembre de 2014). La Inteligencia Artificial augura el fin de la raza humana. *BBC News.* Última visita el 27 de agosto de 2021.

https://www.bbc.com/mundo/ultimas_noticias/2014/12/141202_ultnot_hawking_inteligencia_artificial_riesgo_humanidad_egn

Hidalgo Pérez, M. (25 de noviembre de 2021). La Unesco saca adelante la "declaración universal" de la inteligencia artificial. *El País*.

Hurley, D. (2 de enero de 2018). Can an algorithm tell when kids are in danger? *The New York Times.*

Iglesias Farga, A. (2 de mayo de 2019). De los años 40 a hoy: así ha madurado la Inteligencia Artificial. *El Español*.

India Today Web Desk. (28 de abril de 2018). How Delhi Police tracked 3,000 missing children in only four days with Facial Recognition System. *India Today*.

Kanno-Youngs, Z., & Sanger, D. E. (2019). Border Agency's Images of Travelers Stolen in Hack. *The New York Times*.

Kashmir, H., & J.X. Dance, G. (10 de febrero de 2020). Clearview's Facial Recognition App Is Identifying Child Victims of Abuse. *The New York Times*.

Keller, M. H., & Dance, G. J. (29 de septiembre de 2019). The Internet is overrun with images of child sexual abuse. What went wrong? *The New York Times*.

Koetsier, J. (7 de octubre de 2020). This AI Deepfakes Reality In The Name Of Privacy. *Forbes*.

Kolotushkina, N. (2018). VERIPOL: la herramienta de la Policía para detectar denuncias falsas. *RTVE*. Última visita el 20 de junio de 2021. https://www.rtve.es/noticias/20181102/veripol-herramienta-policia-para-detectar-denuncias-falsas/1831344.shtml

Lane, M. (23 de abril de 2012). ¿Cómo se entera una tienda antes que tus padres de que estás embarazada? *CNN Español*. Última visita el 2 de agosto de 2020. https://cnnespanol.cnn.com/2012/04/23/como-se-entera-una-tienda-antes-que-tus-padres-de-que-estas-embarazada/#0

López de Mántaras Badia, R. (27 de enero de 2016). El legado de un pionero. *El Mundo*.

Markoff, J. (16 de febrero de 2011). Computer Wins on 'Jeopardy!': Trivial, it's not. *The New York Times*.

Marotti, A. (30 de enero de 2020). Facebook may pay Illinois users a couple of hundred dollars each in $550 million privacy settlement. *Chicago Tribune*.

Marotti, A. (24 de julio de 2020). A massive Facebook privacy settlement just got bigger. Illinois users could split $650 million. *Chicago Tribune*.

Masdeu, J. (20 de febrero de 2020). Europa lanza su revolución digital para enfrentarse a EE.UU. y China. *La Vanguardia*.

McDonald, H. (29 de octubre de 2019). AI system for granting UK visas is biased, rights groups claim. *The Guardian*.

McHugh, R., Stulberger, E., & Dienst, J. (16 de abril de 2016). *NBC News*. Última visita el 26 de octubre de 2020. https://www.nbcnews.com/news/us-news/inside-look-system-cut-crime-new-york-75-percent-n557031

Metz, C., & Satariano, A. (6 de febrero de 2020). An algorithm that grants freedom, or takes it away. *The New York Times*.

Mozur, P. (20 de julio de 2020). Beijing wants AI to be Made in China by 2030. *The New York Times*.

Muoio, D. (20 de octubre de 2016). Elon Musk: Tesla not liable for driverless car crashes unless it's design related. *Business Insider*.

Muñoz, M. (9 de junio de 2019). La nueva guerra fría del 5G. *El País*.

Murphy, H. (23 de febrero de 2015). I've Just Seen a (DNA-Generated) Face. *The New York Times*.

Palazuelos, F. (18 de julio de 2017). Elon Musk: "La inteligencia artificial amenaza la existencia de nuestra civilización". *El País*.

Plasencia, A. (12 de julio de 2019). El argumento de la habitación china. *El Español*.

Redacción. (26 de julio de 2011). Figueres estrena un 'ADN sintético' para marcar objetos y localizar al propietario en caso de robo. *La Vanguardia*.

Redacción. (7 de junio de 2012). Schufa will Facebook-Nutzer durchleuchten. *Die Spiegel*.

Redacción. (8 de junio de 2012). Every User Can Decide Alone What Facebook Knows. *Die Spiegel*.

Redacción. (19 de mayo de 2015). Algoritmos contra la violencia machista. *La Vanguardia*.

Redacción. (s.f.) (17 de enero de 2017). Rolls-Royce pagará multas de 760 millones por soborno y corrupción. *ABC*.

Redacción. (30 de octubre de 2017). Sophia, la robot que tiene más derechos que las mujeres en Arabia Saudita. *BBC News*. Última visita el 27 de julio de 2020. https://www.bbc.com/mundo/noticias-41803576

Redacción. (10 de diciembre de 2017). In Your Face: China's all-seeing state. *BBC News*.

Última visita el 28 de junio de 2021. https://www.bbc.com/news/av/world-asia-china-42248056

Redacción. (7 de febrero de 2018). Chinese police spot suspects with surveillance sunglasses. *BBC News*. Última visita el 27 de junio de 2020. https://www.bbc.com/news/world-asia-china-42973456

Redacción. (30 de marzo de 2018). Transparence: Macron veut que l'algorithme de Parcoursup soit rendu public. *20 minutes.*

Redacción. (10 de diciembre de 2018). "Software" berciano para identificación forense. *Diario de León.*

Redacción. (16 de diciembre de 2018). El truco del "turco" que destruyó la reputación de expertos jugadores de ajedrez en la Europa del siglo XVIII. *BBC News*. Última visita el 2 de agosto de 2020. https://www.bbc.com/mundo/noticias-46545215

Redacción. (17 de enero de 2020). La UE plantea prohibir hasta 5 años el reconocimiento facial en lugares públicos para analizar sus riesgos. *Europa Press.*

Redacción. (30 de enero de 2020). Met Police to deploy facial recognition cameras. *BBC News*. Última visita el 12 de noviembre de 2020. https://www.bbc.com/news/uk-51237665

Redacción. (15 de febrero de 2020). Met Police remove 374 names from gangs matrix. *BBC News.* Última visita el 28 de octubre de 2020. https://www.bbc.com/news/uk-england-london-51516812

Redacción. (16 de febrero de 2020). EEUU amenaza a Europa con romper la OTAN: en la guerra contra China por el 5G, o aliados o enemigos. *El Español*.

Redacción. (15 de abril de 2020). Justicia inicia el proceso para la reforma de la LECrim

con la creación de la comisión que elaborará el anteproyecto. *Europa Press.*

Rosenberg, M., Confessore, N., & Cadwalldar, C. (20 de marzo de 2018). La empresa que explotó millones de datos de usuarios de Facebook. *The New York Times*.

Rubin, J. (2010 de agosto de 2010). Stopping Crime Before It Starts. *Los Angeles Times*.

Russon, M. (28 de mayo de 2019). The push towards artificial intelligence in Africa. *BBC News*. Última visita el 18 de septiembre de 2020. https://www.bbc.com/news/business-48139212

Saura, G., & Aragó, L. (6 de diciembre de 2021). Un algoritmo impreciso condiciona la libertad de los presos. *La Vanguardia*.

Saura, G., & Aragó, L. (7 de diciembre de 2021). El algoritmo de prisiones que no rinde cuentas a nadie. *La Vanguardia*.

Sharkey, N. (17 de noviembre de 2018). Mama Mia It's Sophia: A Show Robot Or Dangerous Platform To Mislead? *Forbes*.

Shimabukuro, I. (28 de abril de 2021). Brasileira cria skill da Alexa contra violencia doméstica. *Olhar Digital*. Última visita el 2 de mayo de 2021. http://olhardigital.com.br/en/2021/04/28/pro/brasileira-cria-skill-da-alexa-contra-violencia-domestica/%20%C3%9Altima%20visita%20el%202%20de%20mayo%20de%202021

Smith, M. (22 de junio de 2016). In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures. *The New York Times*.

Smith-Park, L. (24 de agosto de 2014). Voice, words may provide key clues about James Foley's killer. *CNN*. Última visita el 10 de noviembre de 2020. https://edition.cnn.com/2014/08/22/world/europe/british-jihadi-hunt/index.html

Solon, O. (25 de junio de 2020). Facial recognition bill would ban use by federal law enforcement. *NBC News*. Última visita el 9 de noviembre de 2020. https://www.nbcnews.com/tech/security/2-democratic-senators-propose-ban-use-facial-recognition-federal-law-n1232128

Stromboni, C. (21 de enero de 2019). Parcoursup: le Défenseur des droits demande plus de transparence. *Le Monde*.

The Japanese Times. (24 de junio de 2014). *YouTube*. Última visita el 28 de marzo de 2021. https://www.youtube.com/watch?v=Wyl72Re5110

The New York Times Archives. (13 de julio de 1958). Electronic "Brain" teaches Itself. *The New York Times*. Pág. 9.

The Straits Times. (16 de noviembre de 2017). Shanghai aiming to be China's AI hub. *The Straits Times*.

Thomas, D. (17 de julio de 2018). The cameras that know if you're happy - or a threat. *BBC News*. Última visita el 23 de junio de 2021. https://www.bbc.com/news/business-44799239

Tran, M. (12 de febrero de 2021). Deep Blue computer beats world chess champion-archive, 1996. *The Guardian*.

Travieso, J. (17 de abril de 2015). Las consecuencias de mandar a la guerra a "robots asesinos". *El Diario*.

Uberti, D. (25 de junio de 2020). California City Bans Predictive Policing. *The Wall Street Journal*.

Van Dam, A. (19 de noviembre de 2019). Algorithms were supposed to make Virginia judges fairer. What happened was far more complicated. *The Washington Post*.

Vilà Coma, N. (31 de diciembre de 2019). Los robots van a la guerra. *La Vanguardia*.

Wakefield, J. (26 de mayo de 2021). AI emotion-detection software tested on Uyghurs. *BBC News*. Última visita el 13 de julio de 2021. https://www.bbc.com/news/technology-57101248

Wang, K. (7 de diciembre de 2019). China using WeChat for a digital justice system. *Asia Times*.

Weiser, B. (23 de enero de 2017). New York City agrees to settlement over summonses that were dismissed. *The New York Times*.

Williams, T. (12 de junio de 2019). Black People Are Charged at a Higher Rate Than Whites. What if Prosecutors Didn't Know Their Race? *The New York Times*.

Yang, Y. (6 de mayo de 2019). Chinese AR start-up develops smart glasses to help police catch suspects. *South China Morning Post*.

Yonhap. (17 de diciembre de 2019). S. Korea aims to expand prowess in Artificial Intelligence. *The Korea Herald*.

5.3. PUBLICACIONES, NOTAS E INFORMACIONES DE INSTITUCIONES PÚBLICAS

Agencia Española de Protección de Datos (AEPD). (Junio de 2020). Última visita el 3 de marzo de 2021. https://www.aepd.es/sites/default/files/2020-06/nota-equivocos-biometria.pdf

Agencia Europea de Derechos Fundamentales (2018). Última visita el 7 de octubre de 2020. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_fr.pdf+&cd=6&hl=en&ct=clnk&gl=es#23

Agencia Europea de Derechos Fundamentales de la UE. (Junio de 2018). Última visita el 19 de noviembre de 2020. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_en.pdf+&cd=1&hl=en&ct=clnk&gl=es

Agencia Europea de Derechos Fundamentales de la UE. (Junio de 2018). Última visita el 23 de noviembre de 2020. https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/institutions-and-bodies-profiles/fra_es

Agencia Europea de Derechos Fundamentales. (2019). *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights.* Publications office of the European Union.

Ajuntament de Barcelona. Última visita el 23 de julio de 2021. https://ajuntament.barcelona.cat/imi/es/noticia/el-nuevo-sistema-de-reconocimiento-automatico-de-placas-de-matricula-de-la-guardia-urbana_768583+&cd=1&hl=en&ct=clnk&gl=es

Autoridad Catalana de Protección de Datos (APDCAT). Última visita el 29 de septiembre de 2020. https://apdcat.gencat.cat/es/inici/

Baichère, D. (Julio de 2019). *Assemblée Nationale*. Última visita el 6 de noviembre de 2020. https://www2.assemblee-nationale.fr/content/download/179314/1794787/version/2/file/Note%20Reconnaissance%20Faciale%20-%20EN.pdf

Bentley, P., Brundage, M., Häggström, O., & Metzinger, T. (2018). *Should we fear artificial intelligence? In-depth Analysis, 3. The Three Laws of Artificial Intelligence: Dispelling Common Myths.* Parlamento Europeo, STOA-Science and

Technology Options Assessment. Bruselas: European Parliament Research Service.

Comisión Europea. (2008). Última visita el 18 de noviembre de 2020. https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/visa-information-system_en

Comisión Europea. (2016). Última visita el 4 de octubre de 2020. https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52016DC0787&from=ES

Comisión Europea. (10 de mayo de 2017). Última visita el 28 de septiembre de 2020. https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52017DC0228&from=ES

Comisión Europea. (2018). Última visita el 26 de agosto de 2020. https://ec.europa.eu/knowledge4policy/sites/know4pol/files/sweden-ai-strategy-report.pdf

Comisión Europea. (2018). Última visita el 25 de julio de 2020. https://ec.europa.eu/commission/presscorner/detail/es/IP_18_6689

Comisión Europea. (2018). Última visita el 29 de septiembre de 2020. https://ec.europa.eu/info/departments/data-protection-officer_es

Comisión Europea. (2019). Última visita el 25 de marzo de 2020. https://ec.europa.eu/knowledge4policy/ai-watch/estonia-ai-strategy-report_en

Comisión Europea. (2019). Última visita el 23 de mayo de 2020. https://ec.europa.eu/knowledge4policy/ai-watch/czech-republic-ai-strategy-report_en

Comisión Europea. (2019). Última visita el 29 de junio de 2020. https://ec.europa.eu/knowledge4policy/ai-watch/finland-ai-strategy-report_en

Comisión Europea. (2019). Última visita el 29 de septiembre de 2020. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_es#cifras-previstas-para-2025

Comisión Europea. (8 de abril de 2019). Última visita el 3 de octubre de 2020. https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52019DC0168&from=es

Comisión Europea. (2019). Última visita el 6 de octubre de 2020. https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608

Comisión Europea. (2019). Última visita el 25 de octubre de 2020. https://knowledge4policy.ec.europa.eu/ai-watch/denmark-ai-strategy-report_en

Comision Europea. (21 de enero de 2020). *Politico*. Última visita el 6 de mayo de 2020. https://www.politico.eu/wpcontent/uploads/2020/01/SKM_C45820012915530.pdf

Comisión Europea. (2020). *AI Watch Defining Artificial Intelligence.* JRC Technical Reports. Joint Research Center.

Comisión Europea. (19 de febrero de 2020). Última visita el 3 de octubre de 2020. https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2020)64&lang=es

Comisión Europea. (19 de febrero de 2020). Última visita el 7 de octubre de 2020. https://ec.europa.eu/transparency/regdoc/rep/1/2020/ES/COM-2020-64-F1-ES-MAIN-PART-1.PDF

Comisión Europea. (2021). *EU Code of Practice on Disinformation.* Comisión Europea, Bruselas.

Comisión Europea. (2021). *Progress Report: Code of Practice against Disinformation*. Comisión Europea, Bruselas.

Comisión Europea. (21 de abril de 2021). Última visita el 26 de mayo de 2021. https://ec.europa.eu/commission/presscorner/detail/es/ip_21_1682

Comisión Europea. (27 de abril de 2021). Última visita el 22 de junio de 2021. https://knowledge4policy.ec.europa.eu/dataset/ds00008_en

Comisión Europea. *Migration and Home Affairs*. Última visita el 12 de noviembre de 2020. https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/schengen-information-system/sirene-cooperation_en

Comisión Europea. *Migration and Home Affairs*. Última visita el 15 de noviembre de 2020. https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/visa-information-system_en

Comisión Europea. *Migration and Home Affairs*. Última visita el 17 de noviembre de 2020. https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/ees_en

Community Research and Development Information Service (CORDIS). Última visita el 10 de junio de 2021. https://cordis.europa.eu/project/id/700626

Consejo de Europa. (2016). Última visita el 16 de agosto de 2020. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168066cff8

Consejo de Europa. (2017). Última visita el 17 de agosto de 2020. https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10

Consejo de Europa. (2017). Última visita el 17 de agosto de 2020. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016IE5369&from=EN

Consejo de Europa. (2017). Última visita el 18 de agosto de 2020. https://www.coe.int/en/web/freedom-expression/msi-aut

Consejo de Europa. (2017). *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data.*

Consejo de Europa. (2017). Última visita el 17 de noviembre de 2020. https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8.

Consejo de Europa. (19 de octubre de 2017). Última visita el 30 de agosto de 2020. https://pace.coe.int/pdf/af66856450acc0d86abde76d0f976385d75462fbdb5d4c997ffe7b682a054afd/doc.%2014432.pdf

Consejo de Europa. (2018). European Commission for the Efficiency of Justice (CEPEJ). *European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment.*

Consejo de Europa. (2018). Última visita el 31 de agosto de 2020. https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

Consejo de Europa. (2018). Última visita el 5 de septiembre de 2020. https://rm.coe.int/gec-dc-sexism-2018-report-september/16808ec28a

Consejo de Europa. (2018). Última visita el 6 de septiembre de 2020. https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680a2cf96

Consejo de Europa. (2018). Última visita el 7 de septiembre de 2020. https://rm.coe.int/cdcj-2018-5e-technical-study-odr/1680913249

Consejo de Europa. (Junio de 2018). Última visita el 28 de septiembre de 2020. https://rm.coe.int/convention-108-convention-for-the-protection-of-individuals-with-regar/16808b36f1

Consejo de Europa. (3 de diciembre de 2018). Última visita el 5 de octubre de 2020. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

Consejo de Europa. (2019). *Guidelines On Artificial Intelligence And Data Protection*. Última visita el 23 de abril de 2021. https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8

Consejo de Europa. (2019). Última visita el 6 de septiembre de 2020. https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b

Consejo de Europa. (2019). Última visita el 7 de septiembre de 2020. https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

Consejo de Europa. (2019). Última visita el 9 de septiembre de 2020. https://www.coe.int/en/web/artificial-intelligence/cahai

Consejo de Europa. (2019). Última visita el 6 de octubre de 2020. https://www.coe.int/en/web/artificial-intelligence/work-in-progress

Consejo de Europa. (2019). *Unboxing Artificial Intelligence: 10 steps to protect Human Rights.* Última visita el 26 de septiembre de 2020.

https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

Consejo de la Unión Europea. (5 de julio de 2018). Última visita el 13 de noviembre de 2020. https://data.consilium.europa.eu/doc/document/ST-10550-2018-INIT/es/pdf

Consejo de la Unión Europea. (2019). *Conclusions on the significance of 5G to the European Economy and the need to mitigate security risks linked to 5G.*

Consejo de la Union Europea. (14 de mayo de 2019). Última visita el noviembre 16 de noviembre de 2020. https://www.consilium.europa.eu/es/press/press-releases/2019/05/14/interoperability-between-eu-information-systems-council-adopts-regulations/#

Consejo de la Unión Europea. (2020). *Presidency conclusions.The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change.*

Consejo de la Unión Europea. (8 de octubre de 2020). Última visita el 6 de octubre de 2020. https://data.consilium.europa.eu/doc/document/ST-11599-2020-INIT/en/pdf

Courts and Tribunals Judiciary. (17 de enero de 2017). Última visita el 6 de julio de 2021. https://www.judiciary.uk/wp-content/uploads/2017/01/sfo-v-rolls-royce.pdf

Courts and Tribunals Judiciary. Última visita el 21 de mayo de 2021. https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Press-Summary.pdf

Courts and Tribunal Judiciary. (11 de agosto de 2020). Última visita el 12 de noviembre de 2020.

https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Judgment.pdf

Cuerpo Nacional de Policía. Última visita el 23 de noviembre de 2020. https://www.dnielectronico.es/PortalDNIe/PRF1_Cons02.action?pag=REF_2100&id_menu=[6]

EUR-Lex. (3 de junio de 2020). Última visita el 28 de septiembre de 2020. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=LEGISSUM:230105_1

European Travel Information and Authorization System (Etias). Última visita el 22 de noviembre de 2020. https://www.etiasvisa.com/

Europol. (18 de julio de 2019). *Do criminals dream of electric ship? How technology shapes the future of crime and law enforcement.* Pág.12. Última visita el 19 de noviembre de de 2020. https://www.europol.europa.eu/newsroom/news/do-criminals-dream-of-electric-sheep-how-technology-shapes-future-of-crime-and-law-enforcement

Federal Bureau of Investigation (FBI). (16 de mayo de 2011). *YouTube*. Última visita el 20 de mayo de 2021. https://youtu.be/htUgsU4TxXI

G20. (2019). Última visita el 7 de julio de 2020. https://www.mofa.go.jp/files/000486596.pdf

G7. (2019). *Biarritz Strategy for an Open, Free and Secure Digital Transformation*. Última visita el 8 de julio de 2020. https://www.consilium.europa.eu/media/40538/biarritz-strategy-for-an-open-free-and-secure-digital-transformation.pdf

Generalitat de Catalunya. (2015). *Tasa de reincidencia penitenciaria 2014.* Producción propia. Última visita el 21 de febrero de 2021. http://cejfe.gencat.cat/web/.content/home/recerca/cataleg/crono/2015/taxa_reincid

encia_2014/tasa_reincidencia_2014_cast.pdf

Gobierno de Alemania. (2017). Última visita el 6 de junio de 2020. https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connected-driving.pdf?__blob=publicationFile

Gobierno de Alemania. (2018). Última visita el 7 de junio de 2020. https://ec.europa.eu/knowledge4policy/publication/germany-artificial-intelligence-strategy_en

Gobierno de Arabia Saudí. (2021). Última visita el 5 de mayo de 2021. https://ai.sa/Brochure_NSDAI_Summit%20version_EN.pdf

Gobierno de Argentina. (2018). Última visita el 6 de julio de 2020. https://www.argentina.gob.ar/ciencia/plan-nacional-cti/plan-cti

Gobierno de Argentina. (5 de noviembre de 2018). Última visita el 7 de julio de 2020. https://www.argentina.gob.ar/noticias/el-gobierno-presento-la-nueva-agenda-digital-2030

Gobierno de Australia. (2017). Última visita el 23 de junio de 2020. https://www.industry.gov.au/sites/default/files/May%202018/document/pdf/australia-2030-prosperity-through-innovation-full-report.pdf?acsf_files_redirect

Gobierno de Australia. (2018). Última visita el 19 de junio de 2020. https://www.parliament.vic.gov.au/about/news/4029-artificial-intelligence-group-launched

Gobierno de Australia. (2018). Última visita el 21 de junio de 2020. https://www.industry.gov.au/sites/default/files/2018-12/australias-tech-future.pdf

Gobierno de Australia. (2021). Última visita el 27 de julio de 2021. https://www.industry.gov.au/sites/default/files/June%25202021/document/australias-ai-action-plan.pdf

Gobierno de Brasil. (2021). Última visita el 5 de junio de 2021. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_documento_referencia_4-979_2021.pdf

Gobierno de Canadá. (2017). Última visita el 27 de junio de 2020. http://www.jaist.ac.jp/~bao/AI/OtherAIstrategies/Pan-Canadian%20Artificial%20Intelligence%20Strategy.pdf

Gobierno de Canadá. (2018). Última visita el 28 de junio de 2020. https://pm.gc.ca/en/news/backgrounders/2018/12/06/mandate-international-panel-artificial-intelligence

Gobierno de Chile. (2019). Última visita el 6 de julio de 2020. https://www.minciencia.gob.cl/legacy-files/borrador_politica_nacional_de_ia.pdf

Gobierno de Chile. (2019). Última visita el 7 de julio de 2020. file:///Users/borja/Downloads/archivo.pdf

Gobierno de Colombia. (2019). Última visita el 6 de julio de 2020. https://colaboracion.dnp.gov.co/CDT/Conpes/Económicos/3975.pdf

Gobierno de Corea. (2019). Última visita el 19 de junio de 2020. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=5&ved=2ahUKEwiXyLS80obpAhURxoUKHSlKD14QFjAEegQIBhAB&url=https%3A%2F%2Fwww.shinkim.com%2Fattachment%2F16054&usg=AOvVaw2CS0INxDO-J8SeVxc_e7bv

Gobierno de Corea del Sur. (2016). Última visita el 19 de junio de 2020. https://english.msit.go.kr/cms/english/pl/policies2/__icsFiles/afieldfile/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf

Gobierno de Dinamarca. (2018). Última visita el 19 de junio de 2020. https://eng.em.dk/media/10566/digital-growth-strategy-report_uk_web-2.pdf

Gobierno de Dinamarca. (2019). Última visita el 19 de junio de 2020. https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf

Gobierno de EEUU. (2019). Última visita el 2 de julio de 2020. https://www.whitehouse.gov/wp-content/uploads/2019/06/National-AI-Research-and-Development-Strategic-Plan-2019-Update-June-2019.pdf

Gobierno de EEUU. (2020). Última visita el 3 de julio de 2020. https://www.federalregister.gov/documents/2020/01/13/2020-00261/request-for-comments-on-a-draft-memorandum-to-the-heads-of-executive-departments-and-agencies

Gobierno de España. (2007). Última visita el 26 de noviembre de 2020. http://www.interior.gob.es/web/servicios-al-ciudadano/violencia-contra-la-mujer/sistema-viogen

Gobierno de España. (2008). Última visita el 10 de noviembre de 2020. http://www.interior.gob.es/prensa/noticias/-/asset_publisher/GHU8Ap6ztgsg/content/id/9071103

Gobierno de España. (27 de octubre de 2018). Última visita el 18 de junio de 2021. http://www.interior.gob.es/prensa/noticias/-/asset_publisher/GHU8Ap6ztgsg/content/id/9496864

Gobierno de España. (2019). Última visita el 29 de octubre de 2020. https://pnsd.sanidad.gob.es/delegacionGobiernoPNSD/relacionesInternacionales/u

nionEuropea/docs/20190725_InformeDrogasEspana2019_EMCDDA_DGPNSD.pdf

Gobierno de España. (2020). *La Moncloa*. Última visita el 12 de enero de 2021. https://www.lamoncloa.gob.es/presidente/actividades/Paginas/2020/230720-sanchezdigital.aspx

Gobierno de España. (2021). *La Moncloa*. Última visita el 12 de marzo de 2021. https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf

Gobierno de España. (2021). Última visita el 26 de noviembre de 2020. https://violenciagenero.igualdad.gob.es/informacionUtil/recursos/dispositivosControlTelematico/home.htm

Gobierno de España. Ministerio de Ciencia. (2019). Última visita el 10 de marzo de 2020. http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf

Gobierno de España. Ministerio de Defensa. Instituto Español de Estudios Estratégicos. *Documentos de Seguridad y Defensa 79. La Inteligencia Artificial aplicada a la defensa*.

Gobierno de España. Ministerio de Industria. (11 de noviembre de 2017). Nota de prensa.

Gobierno de España. Ministerio de Industria. (s.f.). *Industria Conectada 4.0*. Última visita el 20 de agosto de 2021. https://www.industriaconectada40.gob.es/programas-apoyo/Paginas/programas.aspx

Gobierno de España. Ministerio de Industria. (2015). *Industria Conectada 4.0*. Última visita el 22 de agosto de 2021.

https://www.industriaconectada40.gob.es/SiteCollectionDocuments/informe-industria-conectada40.pdf

Gobierno de España, Ministerio del Interior. (s.f.). Última visita el 15 de noviembre de 2020. http://www.interior.gob.es/en/web/servicios-al-ciudadano/extranjeria/acuerdo-de-schengen/sistema-de-informacion-de-shengen

Gobierno de España. Ministerio de Transformación Digital. (2013). Última visita el 12 de mayo de 2020. https://avancedigital.gob.es/planes-TIC/agenda-digital/DescargasAgendaDigital/Plan-ADpE_Agenda_Digital_para_Espana.pdf

Gobierno de España. Ministerio de Transformación Digital. (Octubre de 2015). Última visita el 15 de mayo de 2020. https://www.plantl.gob.es/tecnologias-lenguaje/PTL/Bibliotecaimpulsotecnologiaslenguaje/Detalle%20del%20Plan/Plan-Impulso-Tecnologias-Lenguaje.pdf

Gobierno de Estados Unidos. (2002). Última visita el 7 de octubre de 2020. https://georgewbush-whitehouse.archives.gov/nsc/nss/2002/

Gobierno de Estados Unidos. (2014). *Big Data: Seizing Opportunities, Preserving Values.*

Gobierno de Estados Unidos. (2017). Última visita el 4 de julio de 2020. https://www.federalregister.gov/documents/2020/01/13/2020-00261/request-for-comments-on-a-draft-memorandum-to-the-heads-of-executive-departments-and-agencies

Gobierno de Estados Unidos. (2017). Última visita el 28 de julio de 2020. https://info.publicintelligence.net/OCIA-ArtificialIntelligence.pdf

Gobierno de Estados Unidos. (2018). Última visita el 29 de junio de 2020. https://www.whitehouse.gov/briefings-statements/president-donald-j-trump-is-accelerating-americas-leadership-in-artificial-intelligence/

Gobierno de Estados Unidos. (2019). Última visita el 28 de junio de 2020. https://www.whitehouse.gov/ai/

Gobierno de Estados Unidos. (2019). Última visita el 5 de julio de 2020. https://www.defense.gov/Explore/News/Article/Article/1755942/dod-unveils-its-artificial-intelligence-strategy/

Gobierno de Estados Unidos. (2019). Última visita el 5 de julio de 2020. https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf

Gobierno de Estados Unidos. (2020). Última visita el 4 de julio de 2020. https://www.whitehouse.gov/wp-content/uploads/2020/02/American-AI-Initiative-One-Year-Annual-Report.pdf

Gobierno de Estados Unidos. (2020). Última visita el 5 de agosto de 2020. https://www.whitehouse.gov/wp-content/uploads/2019/03/ap_21_research-fy2020.pdf

Gobierno de Estados Unidos. (2020). Última visita el 8 de diciembre de 2020. https://www.census.gov/library/visualizations/2020/comm/us-hispanic-population-growth.html

Gobierno de Estados Unidos. (2021). Última visita el 18 de noviembre de 2021. https://www.bop.gov/about/statistics/statistics_inmate_ethnicity.jsp

Gobierno de Estonia. (2017). Última visita el 12 de junio de 2020. https://www.mkm.ee/en/news/estonia-allowing-number-self-driving-cars-streets-starting-today

Gobierno de Estonia. (2019). Última visita el 10 de junio de 2020. https://f98cc689-5814-47ec-86b3-db505a7c3978. filesusr.com/ugd/7df26f_27a618cb80a648c38be427194affa2f3.pdf

Gobierno de Finlandia. (2017). Última visita el 12 de junio de 2020. http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y

Gobierno de Finlandia. (2018). Última visita el 12 de junio de 2020. http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf

Gobierno de Francia. (2018). Última visita el 13 de junio de 2020. https://www.enseignementsup-recherche.gouv.fr/cid136649/la-strategie-nationale-de-recherche-en-intelligence-artificielle.html

Gobierno de Francia. (2018). *AI for Humanity*. Última visita el 13 de junio de 2020. https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

Gobierno de Francia. (2020). Última visita el 13 de junio de 2020. https://www.immigration.interieur.gouv.fr/Europe-et-International/La-circulation-transfrontiere/Le-passage-rapide-aux-frontieres-exterieures-PARAFE

Gobierno de Hungría. (2020). Última visita el 14 de junio de 2020. https://ai-hungary.com/files/e8/dd/e8dd79bd380a40c9890dd2fb01dd771b.pdf

Gobierno de India. (2018). Última visita el 17 de junio de 2020. https://www.niti.gov.in/national-strategy-artificial-intelligence#:~:text=Consequently%2C%20NITI%20Aayog%20released%20India's,AI%20Garage%20of%20the%20World

Gobierno de Italia. (2018). Última visita el 17 de junio de 2020. https://ia.italia.it/assets/librobianco.pdf

Gobierno de Italia. (2018). Última visita el 18 de junio de 2020. https://ia.italia.it/en/ai-in-italy/

Gobierno de Italia. (2020). Última visita el 19 de enero de 2021. https://www.mise.gov.it/images/stories/documenti/Proposte_per_una_Strategia_italiana_AI.pdf

Gobierno de la República Checa. (2019). Última visita el 17 de junio de 2020. https://www.mpo.cz/assets/en/guidepost/for-the-media/press-releases/2019/5/NAIS_eng_web.pdf

Gobierno de Malta. (2019). Última visita el 7 de junio de 2020. https://malta.ai/wp-content/uploads/2019/11/Malta_The_Ultimate_AI_Launchpad_vFinal.pdf

Gobierno de Mexico. (s.f.). Última visita el 6 de julio de 2020. https://www.ia2030.mx

Gobierno de Mexico. (2013). Última visita el 5 de julio de 2020. https://www.gob.mx/mexicodigital

Gobierno de Mexico. (2018). Última visita el 5 de julio de 2020. https://framework-gb.cdn.gob.mx/data/institutos/edn/3dic_publicacion_final_EDN.pdf

Gobierno de Mexico. (2018). Última visita el 5 de julio de 2020. https://www.gob.mx/mexicodigital

Gobierno de Mexico. (2018). Última visita el 6 de julio de 2020. https://www.gob.mx/mexicodigital/articulos/estrategia-de-inteligencia-artificial-mx-2018

Gobierno de Noruega. (2016). Última visita el 23 de junio de 2020. https://www.regjeringen.no/contentassets/07b212c03fee4d0a94234b101c5b8ef0/en-gb/pdfs/digital_agenda_for_norway_in_brief.pdf

Gobierno de Noruega. (2018). Última visita el 23 de junio de 2020. https://digital21.no/wp-content/uploads/2018/09/Digital21_strategi_2018.pdf

Gobierno de Noruega. (2020). Última visita el 23 de junio de 2020. https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=1

Gobierno de Nueva Zelanda. (2017). Última visita el 24 de junio de 2020. https://aiforum.org.nz/about/

Gobierno de Nueva Zelanda. (2018). Última visita el 25 de junio de 2020. https://aiforum.org.nz/wp-content/uploads/2018/07/AI-Report-2018_web-version.pdf

Gobierno de Nueva Zelanda. (2018). Última visita el 25 de junio de 2020. https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf

Gobierno de Países Bajos. (2018). Última visita el 26 de junio de 2020. https://www.government.nl/documents/reports/2018/06/01/dutch-digitalisation-strategy

Gobierno de Países Bajos. (2019). Última visita el 26 de junio de 2020. https://ec.europa.eu/knowledge4policy/ai-watch/netherlands-ai-strategy-report_en

Gobierno de Reino Unido. (2017). Última visita el 27 de junio de 2020. https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk

Gobierno de Reino Unido. (2017). Última visita el 27 de junio de 2020. https://www.gov.uk/government/publications/uk-digital-strategy/uk-digital-strategy

Gobierno de Reino Unido. (2017). Última visita el 28 de junio de 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf

Gobierno de Reino Unido. (2018). Última visita el 27 de junio de 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal__4_.pdf

Gobierno de Reino Unido. (2018). Última visita el 1 de noviembre de 2020. https://www.london.gov.uk/press-releases/mayoral/mayor-publishes-gangs-matrix-review

Gobierno de Reino Unido. (2019). Última visita el 7 de noviembre de 2020. https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/12/27112019-EC-Item-3-Briefing-Note-NDAS-MSV.pdf

Gobierno de Reino Unido. (2020). Última visita el 2 de noviembre de 2020. https://www.london.gov.uk/press-releases/mayoral/mayors-intervention-of-met-gangs-matrix

Gobierno de Reino Unido. (2020). Última visita el 29 de enero de 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/871178/A_guide_to_using_AI_in_the_public_sector__web_version.pdf

Gobierno de Rusia. (2017). Última visita el 15 de junio de 2020. http://www.kremlin.ru/acts/bank/44731/page/2

Gobierno de Rusia. (2018). Última visita el 15 de junio de 2020. http://en.kremlin.ru/events/president/news/57425

Gobierno de Rusia. (2019). Última visita el 16 de junio de 2020. http://www.kremlin.ru/acts/bank/44731

Gobierno de Rusia. (2019). Última visita el 17 de junio de 2020. https://digital.ac.gov.ru/

Gobierno de Singapur. (2019). Última visita el 18 de junio de 2020. https://www.smartnation.gov.sg/docs/default-source/default-document-library/national-ai-strategy-summary.pdf?sfvrsn=55179e0f_4

Gobierno de Singapur. (2019). Última visita el 19 de junio de 2020. https://www.pmo.gov.sg/Newsroom/DPM-Heng-Swee-Keat-at-SFF-X-SWITCH-2019

Gobierno de Singapur. (2020). Última visita el 18 de junio de 2020. https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf

Gobierno de Suecia. (2018). Última visita el 19 de junio de 2020. https://www.government.se/491fa7/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf

Gobierno de Suecia y Consejo Nórdico de Ministros. (2018). Última visita el 9 de julio de 2020. https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf

Gobierno de Turquía. (2021). Última visita el 2 de octubre de de 2021. https://cbddo.gov.tr/SharedFolderServer/Genel/File/TR-NationalAIStrategy2021-2025.pdf

Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects.* National Institute of Standards and Technology, Information Access Division Information Technology Laboratory (NIST). Internal Report.

Hangzhou Internet Court. (s.f.). Última visita el 6 de septiembre de 2021. https://www.netcourt.gov.cn/?lang=En

Illinois General Assembly. Última visita el 8 de junio de 2021. https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57

Instituto Nacional de Estadística. (2020). Última visita el 25 de octubre de 2020. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=estadistica_C&cid=1254736176793&menu=ultiDatos&idp=1254735573206

Interpol. Última visita el 12 de octubre de 2020. https://www.interpol.int/es/Como-trabajamos/Policia-cientifica/Reconocimiento-facial

Interpol. Última visita el 12 de noviembre de 2020. https://www.interpol.int/en/Who-we-are/Member-countries

Interpol. Última visita el 21 de enero de 2021. https://www.interpol.int/es/Como-trabajamos/Policia-cientifica/Huellas-dactilares

Interpol. (2011). Última visita el 12 de agosto de 2021. https://docplayer.es/17495805-Reglamento-de-interpol-sobre-el-tratamiento-de-datos.html

International Orgnanization for Standarization (ISO). Última visita el 12 de noviembre de 2020.
https://www.iso.org/standard/78648.html

Lee, J. (28 de noviembre de 2017). MAS steers debate on ethics of AI, Big Data; Kicks off industry consult. *The Business Times*.

Marciano, M. A., & Adelman, J. D. (2017). *PACE: Probabilistic Assessment for Contributor Estimation-A Machine Learning-based Assessment of the Number of Contributors in DNA Mixtures.* National Institute of Justice.

Montreal Declaration. Última visita el 13 de junio de 2021. https://www.montrealdeclaration-responsibleai.com/the-declaration

National Institute of Justice (NIJ). (2016). Última visita el 10 de marzo de 2021.
https://nij.ojp.gov/funding/awards/2016-dn-bx-0183

National Institute of Standards and Technology. (9 de enero de 2017). Última visita el 5 de mayo de de 2021.
https://www.nist.gov/oles/forensic-database-questioned-documents-table

National Transport Safety Board Office of Highway. (18 de marzo de 2018). Última visita el 4 de octubre de 2020.
https://www.documentcloud.org/documents/6540547-629713.html

OECD. Última visita el 16 de abril de de 2020.
https://oecd.ai

OECD. (2019). Última visita el 10 de junio de 2020.
https://www.oecd.ai/dashboards/policy-initiatives/2019%2Fdata%2FpolicyInitiatives%2F25502

OECD. (2019). Última visita el 18 de junio de 2020.

https://www.oecd-ilibrary.org/sites/eedfee77-en/1/2/5/index.html?itemId=/content/publication/eedfee77-en&_csp_=5c39a73676a331d76fa56f36ff0d4aca&itemIGO=oecd&itemContentType=book

OECD. (2019). Última visita el 21 de junio de 2020. https://oecd.ai/dashboards/policy-initiatives/2019%2Fdata%2FpolicyInitiatives%2F25312

OECD. (2019). Última visita el 22 de junio de 2020. https://oecd.ai/dashboards/policy-initiatives/2019%2Fdata%2FpolicyInitiatives%2F24471

OECD. (2019). Última visita el 23 de junio de 2020. https://oecd.ai/dashboards/policy-initiatives/2019%2Fdata%2FpolicyInitiatives%2F24385

OECD. (2019). Última visita el 6 de julio de 2020. https://www.oecd.org/innovation/oecd-creates-expert-group-to-foster-trust-in-artificial-intelligence.htm

OECD. (2019). Última visita el 7 de julio de 2020. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OECD. (2019). Última visita el 7 de julio de 2020. https://www.oecd.org/going-digital/ai/principles/

OECD. (2019). Última visita el 8 de julio de 2020. https://www.oecd.org/about/secretary-general/artificial-intelligence-g7-summit-france-august-2019.htm

OECD. (2019). Última visita el 8 de julio de 2020.

https://www.oecd-ilibrary.org/sites/eedfee77-en/1/2/5/index.html?itemId=/content/publication/eedfee77-en&_csp_=5c39a73676a331d76fa56f36ff0d4aca&itemIGO=oecd&itemContentType=book#endnotea5z4

Parlamento Europeo. (2019). Última visita el 26 de julio de 2020. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf

Parlamento Europeo. (2020). Última visita el 2 de enero de 2021. https://www.europarl.europa.eu/news/en/press-room/20201208IPR93304/deal-on-an-enhanced-information-system-for-visas-in-the-eu

Parlamento Europeo. (9 de junio de 2020). Última visita el 17 de noviembre de 2020. https://www.europarl.europa.eu/doceo/document/E-9-2020-000173-ASW_EN.html

Parlamento Europeo. (Julio de 2020). Última visita el 26 de agosto de 2020. https://www.europarl.europa.eu/news/en/press-room/20200615IPR81228/parliament-sets-up-special-committees-and-a-permanent-subcommittee

Serious Fraud Office. (10 de abril de 2018). Última visita el 6 de julio de 2021. https://www.sfo.gov.uk/2018/04/10/ai-powered-robo-lawyer-helps-step-up-the-sfos-fight-against-economic-crime/

Surveillance Camera Commissioner. (2018). *Surveillance Camera Commissioner Annual Report 2016/17.* Última visita el 23 de noviembre de 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/672286/CCS207_CCS0118716124-1_Annex_A_-_AR_2017-_web.pdf

Tribunal Supremo chino. (2019). Última visita el 28 de septiembre de 2021.

https://english.court.gov.cn

UNESCO. (2017). Última visita el 7 de julio de 2020. https://unesdoc.unesco.org/ark:/48223/pf0000253952

UNESCO. (2019). Última visita el 18 de septiembre de 2020. https://en.unesco.org/news/forum-benguerir-morocco-examine-challenges-and-opportunities-artificial-intelligence-africa

UNESCO. (Mayo de 2019). *UNESDOC*. Última visita el 3 de octubre de 2020. https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

UNICRI. (2017). Última visita el 8 de julio de 2020. http://www.unicri.it/news/article/2017-09-07_Establishment_of_the_UNICRI

Unión Europea. (2004). Última visita el 28 de septiembre de 2020. https://europa.eu/european-union/about-eu/institutions-bodies/european-data-protection-supervisor_es

Unión Europea. (2018). Última visita el 28 de septiembre de de 2020. https://europa.eu/european-union/about-eu/institutions-bodies/european-data-protection-board_es

Unión Europea. (2019). Última visita el 29 de septiembre de 2020. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_es

World Commission on the Ethics of Scientific Knowdlege and Technology (COMEST). (2017). *Report of COMEST on robotics ethics.*

5.4. OTRAS PUBLICACIONES, NOTAS E INFORMACIONES

About Amazon. (s.f.). Última visita el 6 de noviembre de 2020.

https://www.aboutamazon.com/news/policy-news-views/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition?ots=1&ascsubtag=%5B%5Dvg%5Be%5D21051142%5Br%5Dgoogle.com%5Bt%5Dw%5Bd%5DD

Acemoglu, D., & Restrepo, P. (Diciembre de 2018). *TNIT News*. Última visita el 26 de septiembre de 2020. https://idei.fr/sites/default/files/IDEI/documents/tnit/newsletter/newsletter_tnit_2019.pdf

ACM. (12 de enero de 2017). Última visita el 10 de abril de 2020. http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

ACM Ethics. (2018). Última visita el 15 de abril de 2020. https://ethics.acm.org/2018-code-draft-3/

ACNUR. (Marzo de 2018). Última visita el 26 de septiembre de 2020. https://eacnur.org/blog/derechos-humanos-articulo-1-tc_alt45664n_o_pstn_o_pst/

Advanced Institute for Artificial Intelligence. (2019). Última visita el 12 de abril de 2020. https://drive.google.com/file/d/16x2ziMBEEJUWXmXbKuBSj9mV1VRAKqTX/view

Advancing Pretrial Policy & Research. Última visita el 2 de diciembre de 2020. https://advancingpretrial.org/psa/factors/

Advancing Pretrial Policy & Research. (2020). Última visita el 3 de diciembre de 2020. https://mailchi.mp/7f49d0c94263/our-statement-on-pretrial-justice?e=a01efafabd

Aeropuerto de París. (s.f.). Última visita el 23 de noviembre de 2020. https://www.parisaeroport.fr/en/page-404

Next Generation Artificial Intelligence Research Center. University of Tokio. Última visita el 3 de julio de 2020.
https://www.ai.u-tokyo.ac.jp/ja/activities/the-tokyo-statement

AI Ethics Lab. (2020). *AI Ethics Lab*. Última visita el 4 de abril de 2020. https://aiethicslab.com/big-picture/

AINED. (2018). Última visita el 28 de junio de 2020.
https://www.nwo.nl/en/news-and-events/news/2018/11/companies-and-researchers-publish-roadmap-for-implementing-a-national-ai-strategy.html

Alston, P. (2019). *Brief by the UN Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM c.s./De Staat der Nederlanden (SyRI) before the District Court of The Hague (case number: C/09/550982/HA ZA 18/388).*

Alston, P. (2019). *Brief as Amicus Curiae in the case of NJCM c.s/De Staat der Nederlanden (SyRI) before the District Court of The Hague.* Organización de las Naciones Unidas.

Ambos, K. (28 de mayo de 2020). Última visita el 17 de noviembre de 2020.
https://www.justsecurity.org/70264/the-terrorist-as-a-potentially-dangerous-person-the-german-counterterrorism-regime/

American Civil Liberties Union. (15 de febrero de 2017). Última visita el 8 de junio de 2021.
https://www.aclusandiego.org/sites/default/files/wp-content/uploads/2017/02/2017-02-15-Complaint-FILED.pdf

American Civil Liberties Union. (12 de marzo de 2020). Última visita el 9 de noviembre de 2020.
https://www.aclu.org/aclu-v-dhs-face-recognition-surveillance-complaint

American Civil Liberties Union. (12 de marzo de 2020). Última visita el 9 de noviembre de 2020.
https://www.aclu.org/press-releases/aclu-challenges-dhs-face-recognition-secrecy

Amercial Civil Liberties Union. (28 de mayo de 2020). Última visita el 20 de mayo de 2021.
https://www.aclu.org/press-releases/aclu-sues-clearview-ai

American Civil Liberties Union. (24 de junio de 2020). Última visita el 9 de noviembre de 2020.
https://www.aclu.org/press-releases/man-wrongfully-arrested-because-face-recognition-cant-tell-black-people-apart

American Civil Liberties Union of Pennsylvania. Última visita el febrero de 2020. https://www.aclupa.org/en/about/about-us

Amnistía Internacional. Última visita el 7 de julio de 2021.
https://www.es.amnesty.org/en-que-estamos/temas/vigilancia-masiva/

Amnistía Internacional. Última visita el 28 de julio de 2021.
https://www.es.amnesty.org/en-que-estamos/temas/vigilancia-masiva/

Amnistía Internacional. (2018). *Trapped in the Matrix: Secrecy, stigma, and bias in the Met's Gangs Database.*

Amnistía Internacional. (27 de abril de 2021). Última visita el 12 de junio de 2021. https://www.amnesty.org/en/latest/news/2021/04/russia-police-target-peaceful-protesters-identified-using-facial-recognition-technology/

Amnistía Internacional. (18 de mayo de 2020). Última visita el 30 de octubre de 2020. https://www.amnesty.org.uk/london-trident-gangs-matrix-metropolitan-police

Amnistía Internacional. (20 de mayo de 2019). Última visita el 8 de octubre de 2020.

https://ec.europa.eu/jrc/communities/sites/jrccties/files/expert-meeting-predictive-policing_may2019_report.pdf

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete Problems in AI Safety. Última visita el 3 de noviembre de 2020. https://arxiv.org/pdf/1606.06565.pdf%20http://arxiv.org/abs/1606.06565.pdf

Amoruso, L., Bruno, M., & Dominino, M. (2007). Algunas diferencias entre modelos simbólicos y conexionistas. *XIV Jornadas de Investigación y Tercer Encuentro de Investigadores en Psicología del Mercosur* (pág. 338). Buenos Aires: Facultad de Psicología - Universidad de Buenos Aires.

Andrews, D.A; Bonta, J. Última visita el 2 de diciembre de 2020. https://storefront.mhs.com/collections/lsi-r

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias.* ProPublica.

Apple. Última visita el 26 de marzo de 2020. https://www.apple.com/es/siri/

Armonk, N. (21 de julio de 2010). *IBM*. Última visita el 27 de octubre de de 2020. https://newsroom.ibm.com/2010-07-21-Memphis-Police-Department-Reduces-Crime-Rates-with-IBM-Predictive-Analytics-Software

Arnold Ventures. (15 de octubre de 2015). Última visita el 30 de noviembre de 2020. https://www.arnoldventures.org/newsroom/laura-and-john-arnold-foundation

Asia House. (6 de diciembre de 2018). Última visita el 29 de septiembre de 2020. https://asiahouse.org/news-and-views/kai-fu-lee-age-ai-china-new-opec/

Automatic Sentiment Analysis in the Wild (SEWA). Última visita el 7 de junio de 2021. https://www.sewaproject.eu/

AVPD. Última visita el 29 de junio de de 2020.
https://www.avpd.euskadi.eus/s04-5213/es/

BAAI. (2019). Última visita el 8 de abril de 2020.
https://www.baai.ac.cn/news/beijing-ai-principles-en.html

Barabas, C., & Benjamin, R. (16 de julio de 2019). Última visita el 3 de diciembre de 2020.
https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf

Bennett Moses, S. (9 de junio de 2020). *University of New Wales South*. Última visita el 13 de octubre de 2020.
https://newsroom.unsw.edu.au/news/business-law/predictive-policing-will-you-do-time-crime

Berkman Klein Center & MIT Media Lab. (2018). *Cyber Harvard*. Última visita el 25 de abril de 2020.
https://cyber.harvard.edu/sites/default/files/2018-07/2018-02-12_AIAlgorithmsJusticeOnePager.pdf

Big Brother Watch. (6 de abril de 2018). Última visita el 25 de noviembre de 2020.
https://bigbrotherwatch.org.uk/all-media/police-use-experian-marketing-data-for-ai-custody-decisions

Big Brother Watch (Junio de 2018). Última visita el 23 de julio de 2021. https://bigbrotherwatch.org.uk/wp-content/uploads/2018/07/Big-Brother-Watch-evidence-Policing-for-the-future-inquiry.pdf+&cd=1&hl=en&ct=clnk&gl=es

Boletín Oficial del Estado (BOE). (21 de julio de 2020). Última visita el 13 de octubre de 2020.
https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-8276

Bogus, J. S. (1 de julio de 2019). *American Civil Liberties Union (ACLU).* Última visita el 12 de junio de 2021. https://www.aclu.org/blog/privacy-technology/surveillance-technologies/bogus-aggression-detectors-are-audio-recording

Bontrager, P., Roy, A., Togelius, J., Memon, N., & Ross, A. (2018). DeepMasterPrints: Generating MasterPrints for Dictionary Attacks via Latent Variable Evolution. *IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS).* Los Angeles: IEEE (Institute of Electrical and Electronics Engineers). Biometric Council.

Boulestin, R. (18 de octubre de 2018). *Silicon*. Última visita el 19 de noviembre de 2020. https://www.silicon.fr/renseignement-la-dgsi-semancipe-de-palantir-222231.html

Brighton Park Neighborhood Council, Lucy Parsons Labs, & Organized Communities against Deportations. (Abril de 2021). Última visita el 8 de mayo de 2021. https://www.macarthurjustice.org/wp-content/uploads/2021/05/Motion-for-Leave-to-File-Brief-as-Amici-Curiae-with-Ex.-A-Amicus-Brief-Attached.pdf

Brundage, M., Avin, S., Clark, J., Toner, H., & otros. (Febrero de 2018). Última visita el 19 de septiembre de 2020. https://arxiv.org/pdf/1802.07228.pdf

Buolamwini, J. *MIT Media Lab*. Última visita el 6 de noviembre de 2020. http://gendershades.org/

Burgess, M. (1 de marzo de 2018). *Wired*. Última noticia el 26 de noviembre de 2020. https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit

Burt, T., & Horvitz, E. (1 de septiembre de 2020). *Microsoft.* Última visita el 16 de marzo de 2021.

https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/

Center for Computing History. Última visita el 16 de marzo de 2020. http://www.computinghistory.org.uk/det/43927/UK-Government-Launches-the-Alvey-Programme/

CEOE. (2018). Última visita el 23 de julio de 2020. http://plandigital2025.ceoe.es/wp-content/uploads/2018/10/plan_digital_2025_2018_10_08.pdf

Cheminat, J. (27 de octubre de 2016). *Verizon*. Última visita el 18 de noviembre de 2020. https://www.silicon.fr/big-data-la-dgsi-se-rapproche-de-palantir-161283.html

China Copyright and Media. (14 de junio de 2014). Última visita el 18 de noviembre de de 2020. https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/

Chiu, K. (8 de septiembre de 2020). *Abacus*. Última visita el 18 de noviembre de 2020. https://www.scmp.com/abacus/tech/article/3100516/suzhou-city-takes-page-chinas-social-credit-system-civility-code-rates

Churchill, W. (6 de septiembre de 1943). Última visita el 3 de marzo de 2020. https://winstonchurchill.org/resources/speeches/1941-1945-war-leader/the-price-of-greatness-is-responsibility/

Civio. (2 de julio de 2019). Última visita el 2 de octubre de 2020. https://civio.es/novedades/2019/07/02/que-se-nos-regule-mediante-codigo-fuente-o-algoritmos-secretos-es-algo-que-jamas-debe-permitirse-en-un-estado-social-democratico-y-de-derecho/

Clark, J. (21 de mayo de 2018). *Government News*. Última visita el 19 de junio de 2020. https://www.governmentnews.com.au/government-should-lead-ai-certification-finkel/

Clearview. Última visita el 21 de mayo de de 2021. https://www.clearview.ai/

COPKIT. (2018). Última visita el 28 de enero de 2021. https://copkit.eu

Corbett-Davies, S., & Goel, S. (2018). *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning.* Stanford University.

Crockford, K. (31 de octubre de 2019). *American Civil Liberties Union Massachusetts*. Última visita el 8 de junio de 2021. https://www.aclum.org/en/publications/fbi-tracking-our-faces-secret-were-suing

Cusmariu, A. (4 de enero de 2006). Última visita el 10 de noviembre de 2020. https://www.documentcloud.org/documents/4351987-2006-01-04-Technology-That-Identifies-People-by.html

Darlington, K. (4 de enero de 2017). *BBVA Open Mind*. Última visita el 23 de julio de 2020. https://www.bbvaopenmind.com/tecnologia/inteligencia-artificial/el-comienzo-de-la-era-de-la-inteligencia-artificial/

Data Science Africa. (2013). Última visita el 18 de septiembre de 2020. http://www.datascienceafrica.org

De Larra, M.J. (Enero de 1833). Vuelva usted mañana. *Revista Satírica de Costumbres* (11).

De Paul Velasco, J. (2009). Problemática probatoria de la identificación visual del autor del delito: aportación de la Psicología del testimonio. Consejo General del Poder Judicial. *Psicología del testimonio. Cuadernos Digitales de Formación nº 29*. Págs. 7-18.

Dealer World. (1 de diciembre de 1996). Última visita el 26 de noviembre de 2020. https://www.dealerworld.es/archive/software-de-reconocimiento-de-voz-de-ibm-por-14500-pesetas

Deep Learning INDABA. (2017). Última visita el 18 de septiembre de 2020. https://deeplearningindaba.com/about/our-mission/

Degeling, M., & Berendt, B. (2018). What is wrong with Robocops as consultants? A technology-centric critique of predictive policing. *AI & Society* (33). Pág. 348.

Delgado Romero, C. (enero/febrero de 2020). Comparación Forense del habla: el cambio conceptual. *Revista técnica del Cuerpo Nacional de Policía* (158). Págs. 62-63.

Delgado Romero, C. (2020). Comparación Forense del habla: el cambio conceptual. *Revista técnica del Cuerpo Nacional de Policía*.

Desmarais, S., & Lowder, E. (2019). *Pretrial Risk Assessment Tools.* Safety+Justice Challenge.

Deusto. Última visita el 26 de julio de 2021. https://www.deusto.es/cs/Satellite/deusto/es/universidad-deusto/vive-deusto/humanizar-la-tecnologia-la-universidad-de-deusto-presenta-la-declaracion-de-derechos-humanos-en-entornos-digitales/noticia

D-ID. (2019). Última visita el 3 de febrero de 2020. http://www.deidentification.co/wp-content/uploads/2018/09/White-Paper-GDPR-and-D-ID.pdf

Doshi-Velez, F., & Kortz, M. (2017). *Berkman Klein Center Working Group on Explanation and the Law.* Accountability of AI Under the Law: The Role of Explanation.

DPA. (2014). Última visita el 22 de junio de 2020. https://www.datatilsynet.no/globalassets/global/english/big-data-engelsk-web.pdf

DPA. (2018). Última visita el 17 de julio de 2020. https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

Drexler, K. (2019). *Reframing Superintelligence: Comprehensive AI Services as General Intelligence Technical Report.* Oxford: Future of Humanity Institute, University of Oxford.

Dreyfus, H. (1965). *Alchemy and Artificial Intelligence.* Rand Corporation.

Dyer, C., & Burnett, J. (2014). Exploring Elder Financial Exploitation Victimization: Identifying Unique Risk Profiles and Factors to Enhance Detection, Prevention and Intervention. *Interuniversity Consortium for Political and Social Research.* Texas.

Écoles françaises. Espagne-Portugal (EFEP). (18 de enero de 2018). *EFEP*. Última visita el 6 de mayo de 2021. https://efep.es/es/parcoursup-sustituye-a-apb/

ECYT-AR (21 de enero de 2012). Última visita el 18 de noviembre de 2020. https://cyt-ar.com.ar/cyt-ar/index.php?title=Juan_Vucetich&mobileaction=toggle_view_desktop

Electronic Discovery Reference Mode. Última visita el 6 de agosto de 2021. https://edrm.net/edrm-model/

Ed Markey. (25 de junio de 2020). Última visita el 10 de noviembre de 2020. https://www.markey.senate.gov/news/press-releases/senators-markey-and-merkley-and-reps-jayapal-pressley-to-introduce-legislation-to-ban-government-use-of-facial-recognition-other-biometric-technology

ENFSI & FSAASWG. (2019). 21st International Symposium on the Forensic Sciences of the Australian and New Zeland Forensic Science Society. Pág. 69. Budapest (Hungría).

Equivant. (4 de abril de 2019). Ultima visita el 29 de noviembre de 2020. http://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf

Ericsson. (29 de octubre de 2019). Última visita el 15 de abril de 2020. https://www.ericsson.com/en/blog/2019/10/8-principles-of-ethics-and-AI

Eterno, J., & Silverman, E. (2010). NYPD's Compstat: Compare Statistics or Compose Statistics? *International Journal of Police Science and Management, 12* (3). Págs. 426-449.

Facebook. (2019). *Report on the implementation of the Code of Practice for Disinformation.* Comisión Europea.

FAT ML. (s.f.). Última visita el 12 de agosto de 2020. https://www.fatml.org/resources/principles-for-accountable-algorithms

Fernandez, R. (2019). *Statista*. Última visita el 23 de octubre de 2020. https://es.statista.com/estadisticas/514485/numero-de-condenados-en-espana-por-genero-y-numero-de-delitos/

Ferries, G. (2018). *Home Affairs Select Committee: Policingfor the Future inquiry.* Big Brother Watch.

Flores Gómez, J., & Romero Porro, J. (2020). Ponencia: Por una nueva Ley Orgánica reguladora del Derecho a la Defensa e Intrusismo. *Por una nueva Ley Orgánica reguladora del Derecho a la Defensa* (págs. 3-4). Ilustre Colegio de Abogados de Badajoz.

Florida State University. (2016). *Artificial Inteligence and life in 2030.*

Fox, A. (22 de febrero de 2017). *Govl*. Última visita el 20 de agosto de 2021. https://www.gov1.com/technology/articles/3-tools-helping-law-enforcement-agencies-stop-sex-trafficking-nq91QzbMdg3MasTS/

Fussey, P., & Murray, D. (2019). *Independent Report on the London metropolitan Police Service's Trial of Live Facial Recognition Technology.* University of Essex, Human Rights Centre.

Future of Life Institute. (s.f.). Última visita el 10 de mayo de 2020. https://futureoflife.org/ai-principles/

Future of Life Institute. (s.f.). Última visita el 10 de mayo de 2020. https://futureoflife.org/team/

Future of Life Institute. (2019). Última visita el 17 de junio de 2020. https://futureoflife.org/ai-policy-singapore/

Future of Life Institute. (2019). Última visita el 20 de junio de 2020. https://futureoflife.org/ai-policy-japan/

Future of Life Institute. (2019). Última visita el 22 de junio de 2020. https://futureoflife.org/ai-policy-china/

Future of Life Institute. (2019). Última visita el 27 de junio de 2020. https://futureoflife.org/ai-policy-australia/

Future of Life Institute. (2019). Última visita el 29 de junio de 2020. https://futureoflife.org/ai-policy-united-states/

Future of Privacy Forum. (2018). *Future of Privacy Forum*. Última visita el 2 de abril de 2020. https://fpf.org/wp-content/uploads/2018/06/Beyond-Explainability.pdf

García Márquez, G. (1986). El Cataclismo de Damocles. *Conferencia ofrecida en el 41º aniversario de la bomba de Hiroshima.*

Gascueña, D. (20 de julio de 2017). *BBVA Openmind*. Última visita el 18 de noviembre de 2020. https://www.bbvaopenmind.com/ciencia/grandes-personajes/dactiloscopia-tras-la-huella-del-crimen/

Gerchick, M., DeGroff, S., Beksha, D., Eilers, J., Fukunaga, J., Lemmerman, E., Wisniak, M. (2019). *Risk Assessment Factsheet. Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). Pretrial Release Risk Scale-II. PRRS-II.* Stanford Law School.

González, P., Casas, K., & Mesías, L. (2018). *La transformación policial para el 2030 en América Latina.* Última visita el 12 de febrero de 2021. https://www.thedialogue.org/wp-content/uploads/2018/11/KCasas_TransformacionPolicial_FINAL.pdf

Google. (2018). Última visita el 20 de abril de 2020. https://www.blog.google/technology/ai/ai-principles/

Grand View Research. (2021). *Next Generation Technologies.* Última visita el 12 de junio de 2021. https://www.grandviewresearch.com/industry-analysis/facial-recognition-market

Green, B. (2020). The False Promise of Risk Assessments: Epistemic Reform and the

Limits of Fairness. *Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Pág. 2. Barcelona .

Haas, G., & Poujo, A. (2019). *HAAS Avocats*. Última visita el 2 de octubre de 2020. https://info.haas-avocats.com/droit-digital/parcoursup-le-secret-des-algorithmes-d%C3%A9voil%C3%A9-par-la-justice#_ftn4

Herberg, V., & Lindhoff, A. (24 de enero de 2020). *Frankfurter Rundschau*. Última visita el 19 de noviembre de 2020. https://www.fr.de/politik/hessen-umstrittene-polizei-software-palantir-automatisch-verdaechtig-13454012.html

Herta. (s.f.). Última visita el mayo de 2021. https://hertasecurity.com/advanced-solution-facial-expression-analysis/

Hisham, S. (23 de septiembre de 2019). *Geospatial World*. Última visita el 25 de noviembre de 2020. https://www.geospatialworld.net/blogs/ai-for-policing-future/

Human Rights Watch. (2018). *China: Massive Crackdown in Muslim Region.* Última visita 10 de junio de 2021. https://www.hrw.org/news/2018/09/09/china-massive-crackdown-muslim-region

Human Rights Watch. (2019). *China´s Algorithms of Repression.* Última visita el 28 de junio de 2021. https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass

Human Rights Watch (2020). *Russia expands facial recognition despite privacy concerns*. Última visita el 7 octubre de 2020. https://www.hrw.org/news/2020/10/02/russia-expands-facial-recognition-despite-privacy-concerns

IBM. Última visita el 17 de marzo de de 2020. https://www.ibm.com/watson

IBM. Última visita el 7 de noviembre de 2020. https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice reforms/

IBM. (2018). IBM Watson Group Unveils Cloud-Delivered Watson Services to Transform Industrial R&D, Visualize Big Data Insights and Fuel Analytics Exploration. *Conferencia de prensa de IBM.* Nueva York.

IEE. Última visita el 2 de agosto de 2020. https://ethicsinaction.ieee.org

Intel. (s.f.). Última visita el 2 de mayo de 2020. https://blogs.intel.com/policy/files/2017/10/Intel-Artificial-Intelligence-Public-Policy-White-Paper-2017.pdf

Ionos. (2020). Última visita el 26 de julio de 2020. https://www.ionos.es/digitalguide/online-marketing/vender-en-internet/que-es-la-inteligencia-artificial/

Information Technology Industry Council (ITIC). (s.f.). Última visita el 19 de noviembre de 2020. https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf

Jayadev, R. (2019). Última visita el 27 de noviembre de 2020. https://www.siliconvalleydebug.org/stories/the-future-of-pretrial-justice-is-not-money-bail-or-system-supervision-it-s-freedom-and-community

Jeff Merkley. Última visita el 10 de noviembre de 2020. https://www.merkley.senate.gov/imo/media/doc/20.02.12%20Facial%20Recognition.pdf

Journalism AI. (2019). Última visita el 20 de abril de 2020. https://journalismai.com/2019/05/25/beijing-ai-principles-beijing-academy-of-artificial-intelligence-2019/

JPMorgan Chase Bank, N.A. (s.f.). Última visita el 17 de noviembre de 2020. https://www.chase.com/personal/voice-biometrics

Justia Law. (2018). Última visita el 20 de mayo de 2021. https://cases.justia.com/florida/first-district-court-of-appeal/2018-16-3290.pdf?ts=1545938765

Kang Wei, W., & See Kiat, K. (2014). Última visita el 25 de octubre de 2020. https://dkf1ato8y5dsg.cloudfront.net/uploads/5/82/nec-shaping-tomorrows-safe-cities-today.pdf

Kimaldi. (s.f.). Última visita el 20 de octubre de 2020. https://www.kimaldi.com/blog/biometria/reconocimiento_facial/

Kirkwood, I. (2019). *Betakit*. Última visita el 29 de junio de 2020. https://betakit.com/canadian-quebec-governments-pledge-15-million-to-create-ai-centre-in-montreal/

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). *Inherent Trade-Offs in the Fair Determination of Risk Scores*. Última visita el 14 de marzo de 2021. https://drops.dagstuhl.de/opus/volltexte/2017/8156/pdf/LIPIcs-ITCS-2017-43.pdf

Knorr, E. (13 de abril de 2015). *Infoworld*. Última visita el 19 de agosto de 2021. https://www.infoworld.com/article/2907877/how-paypal-reduces-fraud-with-machine-learning.html

Kofman, A. (2018). *The Intercept*. Última visita el 12 de noviembre de 2020. https://theintercept.com/2018/01/19/voice-recognition-technology-nsa/