

ALGUNOS RETOS ÉTICOS, Y SOBRE TODO JURÍDICOS, DE LA INTELIGENCIA ARTIFICIAL

JOSÉ ANTONIO CASTILLO PARRILLA *

VALLS PRIETO, Javier. *Inteligencia artificial, derechos humanos y bienes jurídicos*. Pamplona: Aranzadi, 2021, 137 pp.

Debo comenzar excusando mi condición de intruso a la hora de realizar esta breve semblanza crítica de la obra del Profesor Javier Valls Prieto. Doblemente intruso, por cuanto soy un civilista que (1) escribe en una revista de Filosofía del Derecho (2) sobre una obra de Derecho Penal. Me parece una circunstancia feliz en la medida en que puede dar cuenta de una creciente interdisciplinarietà en la investigación jurídica¹, pero quiero dejar constancia de ello para situar lo más correctamente posible las expectativas de quien pudiera esperar un juicio crítico sobre los aspectos penales de los que trata la obra². Me centraré en una presentación general de la misma, y me detendré en ciertos aspectos de los que me gustaría debatir (no sólo con su autor).

La obra aborda el fenómeno de la Inteligencia Artificial como un reto que el Derecho aún se encuentra en vías de afrontar, y creo que está en lo cierto. Prueba de ello es no solamente la conocida Propuesta de Reglamento UE de Inteligencia Artificial (AI Act)³, sino los intentos de adaptación de la normativa de producto defectuoso para incluir como tales aquellos basados en IA y la reciente publicación de la Propuesta de Directiva sobre responsabilidad civil de la IA (AI Liability Directive)⁴; el Libro Blanco

* Universidad de Granada. Departamento de Derecho Civil. Plaza de la Universidad, s/n, 18071, Granada (España). Email: castillop@ugr.es

1. Muestra de ello es el Microproyecto “EU Data Market Places”, financiado por la Unidad de Excelencia en Investigación SD2. Sociedad digital: seguridad y protección de derechos, de la Universidad de Granada, que reúne a juristas de casi todas las ramas, informáticos y economistas, en el marco de cuyas actividades se enmarca el presente trabajo, que tengo el honor de dirigir, y en el que participa como miembro del equipo de investigación el Profesor Valls.
2. Debo redirigir, en relación con dicho enfoque, a la recensión del Profesor González Rus: González Rus (2022).
3. Parlamento Europeo y Consejo (2021). Propuesta de Reglamento 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. Bruselas.
4. Parlamento Europeo y Consejo (2022). Propuesta de Directiva sobre adaptación de normas de responsabilidad civil a la inteligencia artificial (AI Liability Directive). Bruselas.

sobre la inteligencia artificial⁵ o el Informe del Parlamento Europeo con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas de 8 de octubre de 2020⁶. Menciono en último lugar el Informe del Parlamento Europeo porque su impulsor, el Europarlamentario Ibán García del Blanco, lamentó en una visita a Granada con ocasión de la presentación de esta obra no haberla tenido a mano mientras preparaban el informe. Creo que estas breves consideraciones dan cuenta de la necesidad y oportunidad de la obra objeto de análisis. Permítanme ahora un breve recorrido por la misma.

La obra se divide en seis bloques si excluimos “Ideas de futuro” y “Conclusiones”. Los bloques son: (1) Definición de IA; (2) Supuestos de riesgo; (3) Principios éticos; (4) De la ética a los derechos fundamentales; (5) Protección de los derechos humanos mediante el sistema penal; y (6) Metodología para comprender los problemas que surgen del uso de los sistemas inteligentes.

El primer bloque pretende situar el marco de la obra, ofreciendo una panorámica de la evolución y sobre todo las diferentes definiciones que se han ido dando sobre la Inteligencia Artificial. La claridad y sencillez en la exposición de ideas de este capítulo es tónica general de toda la obra. El autor repasa brevemente la evolución de las concepciones sobre la IA, y clasifica las herramientas IA en dos grandes grupos: sistemas expertos o *machine learning*, y redes neuronales o *deep learning*. Los sistemas expertos basan su aprendizaje en reglas lógicas (el famoso jugador de ajedrez); mientras que las redes neuronales basan su aprendizaje en datos. Son precisamente estos últimos sistemas los que han permitido el último auge de la inteligencia artificial debido al abaratamiento de los costes de almacenamiento y procesado de datos y a la gran proliferación de todo tipo de datos de la mano de una creciente variedad de dispositivos. Todo ello, a su vez, ha dado lugar a un crecimiento considerable de un sector hasta hace muy poco inexistente como es la economía de los datos, o dicho de otro modo, el aprovechamiento económico de los datos entendidos como activos o materia prima más allá de su consideración como elemento a proteger en tanto que derecho fundamental. Esto último, a su vez, ha dado lugar al surgimiento de mercados de datos, espacios donde paquetes de datos son objeto de intercambio con la finalidad última de favorecer y facilitar su aprovechamiento económico. Todo lo cual ha dado lugar a contundentes

5. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf

6. <https://www.algoritmolegal.com/wp-content/uploads/2021/01/Sobre-aspectos-eticos-de-la-IA-8.10.2020-Ing1%C3%A9s.pdf>

posicionamientos sobre lo peligroso de admitir estos modelos de negocio, que tienen su principal exponente en las obras de Carissa Veliz (2021) y Shoshana Zuboff (2020), o en la Opinión 4/2017 del Supervisor Europeo de Protección de Datos, recomendando suprimir la expresión “datos como contraprestación” de la hoy Directiva 2019/770 (donde, en efecto, dicha expresión no figura).

Como lector, tengo un ligero reproche que hacer a este capítulo, y es que si bien explica la historia y las diferentes concepciones y propuestas de definición de la IA, sus dos grandes campos (*machine learning*; *deep learning*) y los retos a que se enfrenta, no remata con una propuesta propia de concepto de inteligencia artificial. El autor explica a lo largo de la obra que la variedad de herramientas tecnológicas es tan variada que no permite unificar criterios de respuesta sino que prácticamente se tratará de respuestas *ad hoc* para cada caso. Puede que esto justifique su prudencia. A quien escribe, sin embargo, le causa inquietud no disponer de un asidero conceptual (se comparta o no) con el que operar, sino todo un catálogo (bien explicado, eso sí).

El segundo bloque analiza los elementos de la realidad sobre los que ya se están aplicando herramientas de inteligencia artificial y que constituyen a juicio del autor supuestos de riesgo, divididos en (1) supuestos de riesgo relacionados con el uso de la IA por parte de la Administración Pública; y (2) supuestos de riesgo relacionados con el uso de la IA por parte del sector privado. En los primeros, constituyen supuestos de riesgo (1) la Administración de justicia; (2) la prevención y seguridad; (3) los sistemas electorales; (4) la medicina; (5) la gestión de ayudas sociales. En los segundos, corren peligro (1) el mercado de trabajo; (2) la reputación social; (3) la publicidad; (4) el libre mercado; y (5) los seguros y servicios bancarios. Me detengo brevemente en algunos de estos retos. Destacaré algunos sin que deba tomarse como ejercicio de cherry picking o falacia de prueba incompleta: todos los supuestos de riesgo son de interés, pero el espacio de esta reflexión es limitado y es la obra misma la que merece lectura.

Como señala el autor, “en los ámbitos de la justicia civil, mercantil y administrativa el uso de la IA puede ser bastante útil para disminuir la falta de certeza en la resolución que va a dar el juez antes de ir a juicio”, si bien en el ámbito penal todas estas ventajas deben analizarse con mayor cautela en la medida en que se ven afectados derechos fundamentales como la vida y la libertad, y en la medida en que se puede tratar de decisiones discriminatorias por mucho que hayan sido sugeridas (con mayor o menor contundencia) por sistemas IA. Me limito aquí a recordar el contenido del art. 22 RGPD, que prohíbe la toma de decisiones enteramente automatizadas que produzcan efectos jurídicos o afecten significativamente

a los ciudadanos (¿qué mayor efecto jurídico significativo que una pena privativa de libertad?); o el conocido caso COMPAS, que el propio autor explica en este punto. El uso de la IA en procesos electorales, por otra parte, puede conllevar un riesgo de vaciado del derecho fundamental de sufragio (activo y pasivo), como se vio en el caso Cambridge Analytica, explicado en este apartado. Este asunto tuvo un importante eco en España de la mano de la introducción del art. 58 bis de la LOREG con ocasión de la aprobación de la Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales. Dicho artículo, introducido en trámite de enmiendas, pretendía según sus defensores evitar un escándalo similar a Cambridge Analytica en España. Posteriormente suprimido en parte por la Sentencia del TC 76/2019 de 22 de mayo. En lo que se refiere a los riesgos para el sector privado, baste señalar los problemas para el mercado de trabajo en aspectos como la discriminación en la selección de personal que, retroalimentados por unos datos sesgados, relegan a peores puestos (o directamente no contratan) a ciertos grupos de población. El autor señala el caso de Roland Behn, que se percató de que el uso de datos médicos por parte de los sistemas de selección de personal discriminaban a su hijo por padecer un síndrome de bipolaridad; pero podemos recordar otros como el sistema que Amazon tuvo que retirar en 2018 por discriminar a mujeres debido, entre otros aspectos, a que los datos de que se nutría eran los de los perfiles de solicitantes de empleo en Amazon los últimos 10 años⁷.

Los dos siguientes capítulos deben verse, a mi juicio, como dos caras de una moneda. En el tercero se da cuenta de las distintas cartas de principios éticos, y grupos de trabajo que han incidido o que han tratado sobre la Inteligencia Artificial: Instituto Future of Life (2017); Declaración sobre inteligencia artificial, robótica y sistemas autónomos del Grupo europeo sobre ética en ciencia y nuevas tecnologías de la Comisión Europea (2018); HLEG sobre inteligencia artificial (2018); Declaración de Montreal (2018); Guía del Instituto de ingenieros electrónicos y eléctricos (2019); OCDE; Servicio del Gobierno digital y Oficina para la IA del Reino Unido (2019); Principios de la IA de Pekín (2019); Primer borrador de la recomendación sobre la ética de la IA de la UNESCO (2020); y Carta de derechos digitales de España (2021).

En “la otra cara” (Capítulo IV, De la ética a los Derechos Fundamentales) se desarrollan los límites y sobre todo las potencialidades de la ética de cara al afrontamiento jurídico del fenómeno de la Inteligencia Artificial. En palabras del autor, “no es la intención de la ética sustituir al Derecho (...) la

7. https://elpais.com/tecnologia/2018/10/11/actualidad/1539278884_487716.html

reflexión ética puede ayudarnos a comprender el modo en que el desarrollo, despliegue y utilización de la inteligencia artificial puede afectar a los derechos fundamentales y contribuir a ofrecer orientaciones más detalladas de cara a tratar de identificar aquello que debemos hacer ya que el derecho no llega a completar los principios morales y éticos”. Se trata de un planteamiento digno de análisis si tenemos en cuenta la creciente popularidad de la ética de la IA; popularidad que obedece, entre otras muchas razones, a la mayor comodidad que ofrece un elenco de orientaciones cuyo incumplimiento no es objeto de sanción para quien desarrolla estas herramientas.

El último gran capítulo es el que incide, después de una delimitación precisa del marco de debate, en los bienes jurídicos que pudieran verse afectados por la utilización de herramientas IA: privacidad; igualdad/no discriminación; dignidad; libertad/autonomía; libre pensamiento y derechos de reunión; vida e integridad física; bienes económicos (ej, derechos de los consumidores o libre competencia); y administración pública. En coherencia con lo advertido al inicio, no valoraré de forma crítica este apartado puesto que no tengo los elementos necesarios para ello.

Sí quisiera parar, no obstante, en la distinción con que comienza el capítulo entre bienes jurídicos individuales y colectivos y la consideración de la privacidad como un bien colectivo. Mientras que los bienes jurídicos individuales sirven a los intereses de una persona o grupo, los bienes jurídicos colectivos protegen al conjunto social como colectivo. Esto incide en la divisibilidad del bien jurídico a proteger en cada caso: divisible en los colectivos; indivisible en los individuales. La privacidad, en los últimos años, se ha transformado en un bien colectivo⁸, más parecido ya al medio ambiente que a los derechos de no injerencia del Estado surgidos de la Revolución Francesa y de los que precisamente la intimidad es uno de los principales exponentes. ¿Cómo hemos llegado a esto? Una parte se explica a través del fenómeno conocido como privacidad de grupo (*vid.* Taylor, Vander Sloot y Floridi, 2016), y que el autor explica del siguiente modo: “en la actualidad con el uso de grandes cantidades de datos, los datos ya no sólo afectan a las personas individuales, mis datos, mi información personal, sino que al compartir la información afecta a más personas por el análisis que se puede hacer de ellos”. Esta ingente cantidad de datos disponible hoy para su análisis existe no sólo debido a una cuestión puramente tecnológica (mayor cantidad de dispositivos y tiempo de uso; mayor generación de

8. En esto coincido tanto con el autor como con Carissa Veliz, a la que cita (p. 112 de la obra).

datos), sino a una suerte de “tiranía del consentimiento”⁹ o contaminación de consentimientos si queremos continuar con el símil del medio ambiente: una gran cantidad de personas han aceptado de forma irreflexiva que webs y terceros analicen y procesen sus datos (ej, aceptando cookies para poder leer una noticia) y ello ha generado importantes caladeros de datos legalmente utilizables con base en consentimientos irreflexivos de sus titulares. Volviendo a la idea del bien jurídico colectivo, las reflexiones del Profesor Valls en este punto son un buen punto de partida de cara a la necesidad de repensar el funcionamiento de un derecho fundamental (intimidad, privacidad, protección de datos; entendámoslos conjuntamente a estos efectos) que ha pasado de ser fundamentalmente un derecho de no injerencia a ser algo parecido a un “derecho ambiental”. Queda mucho por desarrollar en relación con la creación de “zonas de bajas emisiones contaminantes de consentimientos al tratamiento de datos”, pero es un buen comienzo.

En estas últimas líneas y al hilo del último de los capítulos (Metodología para comprender los problemas jurídicos que surgen del uso de los sistemas inteligentes), me permito compartir una inquietud acerca de la lógica preventiva o anticipatoria de las normas que se están desarrollando en torno a la inteligencia artificial (y no sólo). Refiere el autor (p. 107) la “ética anticipatoria” de Brey como método para la evaluación de la tecnología. Ésta se divide en tres fases de estudio: nivel de evaluación de la tecnología (donde se define una tecnología concreta); nivel de artefacto (configuración física en un entorno concreto para producir un desarrollo deseado); y nivel de aplicación (problemas derivados de su uso). En cuanto al uso inadecuado, a su vez, propone Valls distinguir tres momentos: en la recolección de datos (aspecto en que pretende incidir el RGPD), en su procesado (para ello es de gran importancia la explicabilidad de la IA, que se propone como un nuevo bien jurídico protegido en otro punto de la obra) y en el resultado final (ej, externalidades negativas como pueden ser resultados discriminatorios).

Se destaca al principio de este capítulo que el oscurantismo de la black box hace imposible establecer nexo causal entre acciones y daños cuando interviene la IA. Esta situación se pretende evitar realizando una evaluación previa en dos momentos cruciales: antes de empezar a diseñar el artefacto y antes de su lanzamiento al público en general. Esta perspectiva gobierna, a mi juicio, la Propuesta de Reglamento de IA. No se trata de un mal enfoque en la medida en que el cumplimiento normativo de todas las cautelas previas establecidas por las normas ya vigentes y las que están siendo objeto de debate hoy puede evitar a futuro muchos de los problemas y situaciones de

9. La expresión es de Elena Gil González (2022).

riesgo que se han señalado. Ahora bien, no debe perderse de vista la lógica reactiva o compensatoria y por ello es una buena noticia la publicación de proyectos normativos como los destacados al inicio sobre responsabilidad civil de la IA o reforma de la normativa de productos defectuosos para incluir a herramientas IA o basadas en IA. Tomando palabras de un Registrador de la Propiedad, Joaquín Delgado Ramos, debemos tener cuidado de enfocarnos sólo en la transparencia (aquí, ética anticipatoria, seguridad y privacidad por diseño y por defecto, etiquetado IA y *compliance*), no sea que convirtamos en transparente el abuso y no podamos luchar contra éste.

En definitiva, la obra del Profesor Valls aborda una temática necesaria y oportuna en el momento actual; y lo hace con gran coherencia interna, y claridad y sencillez en los planteamientos. Todo ello hace muy fácil su lectura, amén de lo sugerente de ciertos planteamientos de los que he pretendido dar cuenta en estas reflexiones.

REFERENCIAS BIBLIOGRÁFICAS

- Gil González, E. (2022). *El interés legítimo en el tratamiento de datos personales*. Madrid: Wolters Kluwer.
- González Rus, J. J. (2022). Recensión del libro de Javier Valls Prieto, Inteligencia artificial, Derechos Humanos y bienes jurídicos. *Revista Electrónica de Ciencia Penal y Criminología*, 24-4r2, 1-10.
- Taylor, L.; Van Der Sloot, B.; Floridi, L. (2016). *Group Privacy: the Challenges of New Data Technologies*. Nueva York: Springer
- Vélez, C. (2021). *Privacidad es poder. Datos, vigilancia y libertad en la era digital*. Barcelona: Debate.
- Zuboff, S. (2020). *La era del capitalismo de la vigilancia: la lucha por un futuro humano frente a las nuevas fronteras del poder*. Barcelona: Paidós Ibérica

