

Inteligencia artificial y derecho: entre el mito y la realidad

«La destrucción algorítmica de la humanidad»

Elisa Simó Soler

Contratada Predoctoral FPU (FPU 17/04436)

Elisa.simo@uv.es

Universitat de València

Paolo Rosso

Profesor catedrático

proso@dsic.upv.es

Universitat Politècnica de València

Diario La Ley, Nº 9982, Sección Tribuna, 4 de Enero de 2022, Wolters Kluwer

ÍNDICE

[Inteligencia artificial y derecho entre el mito y la realidad](#)

[I. ¿De qué hablamos cuando hablamos de Inteligencia Artificial?](#)

[II. Desmitificando la Inteligencia Artificial](#)

[III. Lejos del abracadabra](#)

[IV. Efectos \(anti\)discriminatorios](#)

[V. El dilema de la objetividad de la máquina](#)

[VI. Blanco sobre negro: la necesidad de abrir la caja negra](#)

[VII. Soberanía en la era digital](#)

[VIII. La sustitución algorítmica en la Administración de Justicia](#)

[IX. Conclusiones: ¿mito o realidad?](#)

[X. Bibliografía](#)

Normativa comentada

Constitución Española de 27 Dic. 1978

TÍTULO VI. Del Poder Judicial

Artículo 117

Artículo 120

LO 6/1985 de 1 Jul. (Poder Judicial)

TÍTULO PRELIMINAR. Del Poder Judicial y del ejercicio de la potestad jurisdiccional

Artículo 2.

RD 24 Jul. 1889 (Código Civil)

TÍTULO PRELIMINAR. De las normas jurídicas, su aplicación y eficacia

CAPÍTULO II. APLICACIÓN DE LAS NORMAS JURÍDICAS

Artículo 3

Comentarios

Resumen

Los avances en Inteligencia Artificial y su conexión con la Administración de Justicia obligan a realizar un análisis pausado sobre las implicaciones de la progresiva incorporación de sistemas automatizados en el ámbito del reconocimiento de derechos. Sin unos fundamentos teóricos sólidos para comprender adecuadamente la nueva era de la información es posible incurrir en posicionamientos sesgados que impidan reconocer la contribución que la ciencia y la tecnología pueden ofrecer en favor de la protección y la garantía de los derechos.

Palabras clave

Inteligencia Artificial, Justicia, mitos, derechos, rigurosidad.

Abstract

Developments in Artificial Intelligence and their connection with the Administration of Justice make it necessary to carry out a thorough analysis of the implications of the progressive incorporation of automated systems in the sphere of the recognition of rights. Without solid theoretical foundations to properly understand the new information age, it is possible to incur biased positions that prevent us from recognising the contribution that science and technology can offer in favour of the protection and guarantee of rights.

Keywords

Artificial Intelligence, Justice, myths, rights, accuracy.

I. ¿De qué hablamos cuando hablamos de Inteligencia Artificial?

Actualmente resulta complicado mantenerse al margen de los avances en Inteligencia Artificial (IA). Cada vez más las personas han desvinculado la IA del paradigma de la ciencia ficción para incorporarla a su realidad diaria, los medios de comunicación publican noticias sobre los progresos realizados en sectores como la salud, las finanzas, el transporte o la comunicación y en nuestra cotidianidad convivimos con esta tecnología, aunque no siempre lo sepamos identificar. Dada la omnipresencia, consciente o no, de la IA es importante conceptualizarla y definir sus contornos. Por IA se entiende «la capacidad general de una máquina para replicar de forma independiente los procesos intelectuales típicos de la cognición humana al decidir una acción en respuesta a su entorno percibido» (1) . El propósito es reproducir algunos comportamientos intelectuales humanos para resolver problemas, tomar decisiones y aprender del entorno.

La IA se clasifica en débil y fuerte. A través de la primera se considera que el comportamiento humano puede ser modelizado y utilizado por máquinas para resolver problemas más o menos complejos. Los algoritmos realizan tareas específicas como respuesta a su programación, pero sin comprender ni deducir el significado de la orden que se les ha enviado como es el caso de Siri o Alexa. Por su parte, la IA fuerte busca crear sistemas con conciencia humana y emociones. Se apuesta porque la IA es capaz de aprender como lo hace una criatura y de pensar, pero mucho más rápido debido a su potencia de cálculo y capacidad de procesamiento (2) .

Por su parte el Machine Learning (ML) es un subcampo de la IA, una de las estrategias que puede usar la IA en su objetivo de imitar la cognición humana. Supone el desarrollo de algoritmos que analizan información datificada, reconocen patrones y aprenden de los datos en un proceso llamado «entrenamiento» para proporcionar apoyo a la toma de decisiones. Por tanto, se trata de un proceso automatizado que permite encontrar relaciones entre los datos y elaborar predicciones. Los procesos de aprendizaje pueden clasificarse en tres tipos: supervisado, no supervisado y aprendizaje por refuerzo.

En el aprendizaje supervisado los algoritmos necesitan ayuda externa para realizar las tareas de predicción o clasificación. Se proporciona un conjunto de datos de entrada y de salida etiquetados para que un modelo puede aprender algún tipo de patrón que permita predecir o clasificar la variable de salida correctamente. En el caso del aprendizaje no supervisado, a diferencia del anterior, no hay etiquetado. Se proporciona un conjunto de datos de entrada sin clasificar y se le pide a la máquina que busque patrones subyacentes y prediga el resultado. Cuando se introducen nuevos datos, se emplean las características aprendidas previamente para reconocer la clase de los datos. Se utiliza principalmente para la agrupación y la reducción de características. Por último, en el aprendizaje por refuerzo, los algoritmos aprenden a reaccionar a un entorno por sí mismos, a través de un proceso de recompensa acumulada en casos de éxito. Se busca un equilibrio entre la explotación (la maximización de la recompensa) y la exploración (la búsqueda de mejores resultados). Este tipo de aprendizaje se encuentra en el ámbito de la robótica y la industria de los videojuegos (3) .

II. Desmitificando la Inteligencia Artificial

Es innegable la materialización del binomio IA y Derecho. Esta relación surge por la necesidad de regular el desarrollo de la IA respetando los derechos fundamentales y por la entrada de los sistemas de IA en el ámbito de la Justicia. Su grado de implementación depende del nivel de conocimiento sobre estos sistemas y de la percepción de justicia que susciten. Por un lado, es fundamental conocer, de forma rigurosa, los resultados que pueden ofrecer para evitar posicionamientos extremos con tendencias a demonizar o divinizar su uso. Por otro lado, no es suficiente con que la IA sea justa, también tiene que parecerlo. Como ocurre con las leyes, las aplicaciones de IA tienen que convencer a la sociedad de sus virtudes y exponer nítidamente sus defectos. No obstante, resulta conveniente realizar una primera matización. Cuando se habla de IA no se debe definir como un ente suprahumano, monolítico y abstracto. La realidad de la IA es, hoy en día, mucho menos romántica. Se trata de modelos estadísticos complejos capaces de autoajustarse según reciben nueva información. Detrás de cada algoritmo utilizado para crear un modelo etiquetable como IA, hay un conjunto de personas expertas (no solo técnicas) con diversas funciones asignadas. De esta forma, la IA funcionará dentro de los límites que como sociedad fijemos, contando siempre con la posibilidad de pulsar el botón de apagado.

Con la intención de cumplir con ese canon de precisión sin idealización, en este texto se toman una serie de ideas fuerza que se han extendido en el abordaje de la IA desde una perspectiva jurídica con el objetivo de desclasificar determinados preconceptos en torno a la naturaleza de los sistemas de IA y sus consecuencias.

III. Lejos del *abracadabra*

Como advierte WORTHY CAMPBELL, la IA no es una varita mágica (4) , es un tipo de tecnología con muchas capacidades, pero más limitaciones. Por ello no se debe depositar esperanzas inalcanzables en las aplicaciones de IA ni suponer que los resultados obtenidos son fruto de algún fenómeno que escapa al entendimiento humano. En este sentido, tampoco se debe confundir la predicción, propia de los modelos de ML, con una suerte de futurología. Predecir, en una de las acepciones previstas por la RAE, es el anuncio con conocimiento fundado de algo que ha de suceder. Predecir es pronosticar con base en el conocimiento siguiendo una metodología, lo cual, constituye el ejercicio principal de la ciencia y la ingeniería. Los cálculos de probabilidad en áreas como la biología, la astrofísica, la medicina o la química están a la orden del día, sea para diagnosticar una enfermedad, estimar la *ratio* de contagios de un virus o acotar perimetralmente el impacto de un meteorito. Incluso para BOSTROM y YUDKOWSKY una de las funciones más importantes del sistema legal es ser predecible de forma que la función del sistema legal no consiste necesariamente en optimizar la sociedad, sino en proporcionar un entorno predecible dentro del cual la ciudadanía puede optimizar sus propias vidas (5) .

Utilizar técnicas de ML es, al fin y al cabo, analizar datos y analizar datos no puede ir en contra de comprender la realidad. Se establecen unas premisas, se estima el margen de error y se confirman o refutan las hipótesis de partida. De este modo, pese a que hay distintos niveles de supervisión de los modelos de ML, es un ejercicio en todo caso controlado. Además, si en el ámbito de la investigación se considera el impacto de estos modelos atendiendo a la calidad de los estudios empíricos, la posibilidad de enfrentarse a la resolución de un problema sumando más análisis (por la mayor cantidad de datos y la mayor capacidad de procesamiento) es un valor agregado.

En consecuencia, habrá que pensar la IA como un progreso más de la ciencia y en todo caso cuestionar los potenciales usos que se propongan teniendo en cuenta su relación con los derechos humanos, si sirven para protegerlos o los vulneran.

IV. Efectos (anti)discriminatorios

En este nuevo contexto de digitalización, una de las principales limitaciones y riesgos señalados es la producción o perpetuación de sesgos con el uso de IA. La discriminación algorítmica, por su impacto negativo en los derechos de las personas, es una temática ampliamente estudiada (6) . Asumiendo que se trata de una cuestión en boga, cabe realizar una serie de consideraciones.

En cuanto a la procedencia del sesgo, se han identificado tres componentes principales generadores en la discriminación algorítmica: la configuración de una base de datos que recoja las desigualdades existentes en la sociedad, la selección de unas variables de entrenamiento con potenciales efectos discriminadores y la homogeneidad de los equipos de trabajo, generalmente formados por hombres blancos heterosexuales de clase media del sector tecnológico (7) . Sin embargo, estas tres causas no señalan directamente a los modelos de ML como el origen de los sesgos sino al factor humano. En ese caso, la enmienda al empleo de IA debería formularse

atendiendo a esta intervención humana (sea de la mano de generadores de datos, desarrolladores de las aplicaciones e incluso usuarios). En concreto, la responsabilidad de quienes programan, aunque pudiera darse esa mayoría masculina, se reduce al mínimo. Quien configura el *corpus* de datos (selecciona y anota los datos, las métricas y el modo de clasificación) es una persona experta en la materia objeto de estudio. El ingeniero elige el tipo de algoritmo, la arquitectura del modelo y define los parámetros de ajuste (trabaja con una secuencia anidada de fórmulas que procesan y/o clasifican datos según ciertos criterios), pero no interviene en la creación de la base de datos. En consecuencia, sería conveniente reducir el nivel de culpabilidad asociado al algoritmo y a los programadores, así como descartar la inevitabilidad del sesgo estudiando cómo generar bases de datos, cómo escoger variables y cómo incorporar personas expertas para evitar efectos discriminatorios.

La clave está en no esforzarse en exclusiva por detectar los efectos perniciosos de la IA sino dedicar esfuerzos para poner su potencial al servicio del bien común

Respecto a la aplicación de IA en el ámbito del Derecho y la menor parcialidad de las personas en comparación con los algoritmos, hay que partir del cuestionamiento de la supuesta objetividad, neutralidad y racionalidad de las normas cuando han sido formuladas y aplicadas desde el masculino, singular o plural, sin amparar otras identidades que conforman la realidad social. Asimismo, se debate sobre la objetividad de quienes representan al tercer poder del Estado. Como asevera VIVÉS ANTÓN, la función jurisdiccional no se identifica con silogismos ni automatismos. Los jueces no son seres que aplican de forma mecánica el ordenamiento jurídico ni se

limitan a pronunciar las palabras de la ley (8) . Los jueces son personas y esa nota de humanidad conlleva la necesidad de incorporar medidas para proteger la garantía de imparcialidad. Así, en el momento en el que un juez se retira la toga «solo queda un individuo más, con afectos, carencias, angustias, pensamientos propios y una vasta lista de características personales» (9) . En este sentido, son varios los estudios que señalan diferentes condicionantes, tanto estructurales (condiciones políticas, sociales o económicas) como contextuales al supuesto de hecho objeto de enjuiciamiento (el resultado de un partido de fútbol, la fatiga por el número de decisiones adoptadas en un día, la hora del juicio, el tipo de noticias en prensa, entre otras) que influyen en el veredicto final de un caso (10) . Siendo así, cabría plantearse si es válida la afirmación de que, siempre y en todo caso, un algoritmo va a actuar más sesgado que una persona. Junto a esta idea, podría valorarse la posibilidad de atribuir a los sistemas de IA una finalidad antidiscriminatoria constituyendo un instrumento para dismantelar sesgos y prejuicios humanos presentes en diferentes ámbitos como, por ejemplo, los mensajes misóginos y sexistas en las redes sociales (11) . Podría emplearse esta tecnología en proyectos que supongan una mejora de las condiciones de vida de colectivos históricamente oprimidos. No esforzarse en exclusiva por detectar los efectos perniciosos de la IA sino dedicar esfuerzos para poner su potencial al servicio del bien común.

V. El dilema de la objetividad de la máquina

Como se ha anticipado, pese a la identificada técnica de «copiar y pegar» en la redacción de sentencias (12) , se descarta la posibilidad de una aplicación robotizada de las leyes por parte de los jueces. Ello conlleva la denuncia de una deriva deshumanizadora de la justicia cuando la toma de decisiones se basa en cifras sin atender al contexto, a las características del asunto o a las particularidades de las partes implicadas (13) . Este fenómeno de datificación de la justicia obligaría a redefinir el modo en que se interpretan y aplican las normas al desnaturalizar el codiciado principio de inmediación, garante de la correcta valoración de la prueba. Según BATELLI, la diferencia entre un juez robot y un juez humano es la capacidad del segundo de relacionar el texto con el contexto. Es decir, conocer los antecedentes históricos y legislativos, atender a la realidad social del momento en que las leyes tienen que ser aplicadas para evitar obtener fallos absurdos siguiendo un criterio estrictamente literal, tal y como prevé el artículo 3.1 del Código Civil (LA LEY 1/1889) (14) . No obstante, esta distinción es probable que desaparezca ante la posibilidad de reactualizar los datos a las condiciones materiales y con los avances respecto a la ampliación de los inputs y el aumento de la capacidad de procesamiento de forma que cualquier información con respecto al contexto podrá tenerse también en cuenta.

Para algunos autores el efecto no solo es deshumanizador sino también deslegitimador. Tratándose de uno de los poderes del Estado, delegar la resolución de conflictos en algoritmos puede suponer una pérdida de legitimidad del sistema en el que operan (15) . De este modo, determinados desarrollos tecnológicos disminuyen la calidad del sistema democrático y empobrecen al conjunto de la sociedad al despojar a los jueces de esa función genuina de

juzgar y hacer ejecutar lo juzgado desde la creatividad y la adecuación al caso concreto.

Aunque el carácter objetivo de las decisiones respaldadas en datos podría ser leído en términos positivos al ofrecer rigurosidad y una fundamentación corroborable, actualmente prima la desconfianza. Quienes dudan de la transición hacia la digitalización consideran que cuanto menor es la interacción humana, mayores son las dificultades para explicar sus resultados, reduciendo las posibilidades de su uso. No obstante, cabe reflexionar sobre qué habría de humano en dejar que juzgue una persona con más error que una máquina. Podría ser el caso de las víctimas de violencia de género que denuncian haber sufrido victimización secundaria en su paso por los tribunales. Quizá, para evitar ese trato revictimizador fruto del uso de estereotipos de género y de la ausencia de formación especializada, serían proclives a beneficiarse de una digitalización de la justicia. Cada vez más surgen nuevas iniciativas para crear sistemas de aprendizaje automático cuyos procesos sean comprensibles además de técnicas que faciliten la explicabilidad de los mismos. El objetivo es ayudarnos a comprender el desempeño del modelo para determinar si el resultado es preciso e imparcial y, en definitiva, para otorgar o no nuestra confianza a los sistemas IA.

VI. Blanco sobre negro: la necesidad de abrir la caja negra

El problema de las cajas negras es recurrente en Derecho. Se dice que un sistema es una caja negra si el modelo del sistema no se conoce o es demasiado complejo para razonar explícitamente sobre él (16) . El desconocimiento sobre el proceso que conduce a un determinado resultado desincentiva, incluso veta, la posibilidad de utilizar sistemas de IA en Derecho. El artículo 120.3 de la Constitución Española (LA LEY 2500/1978) impone la obligación de motivar las sentencias. La adopción de decisiones judiciales sin un seguimiento sobre la generación y manipulación de los datos, desconociendo el funcionamiento interno, conllevaría el incumplimiento de dicho deber. Para el Derecho, no es suficiente con que las predicciones digan que algo es cierto, hacen falta explicaciones que ilustren por qué ese algo es cierto (17) . Para lograr la trazabilidad de los resultados se han adoptado algunas soluciones parciales como la explicabilidad y la transparencia algorítmica (18) para promover la confianza de la ciudadanía, aunque dichas propuestas no están exentas de controversias (19) .

Sin embargo, en el sistema actual de justicia «analógica», ¿se comprenden con exactitud todas las decisiones? Es decir, ¿se desvela cuál ha sido el proceso cognitivo por el cual se opta por una decisión y no por otra? Sabiendo que en el proceso de toma de decisiones intervienen sesgos cognitivos (20) , ¿cómo de accesible y decodificable es la mente de una persona para revertir las reglas de su circuito cognitivo o cuantificar sus patrones discriminatorios? Siguiendo la postura de BONET NAVARRO, del mismo modo que se alerta sobre los riesgos de las cajas negras en los algoritmos habría que interrogarse, para salvaguardar la independencia y la imparcialidad del poder judicial, sobre los caminos (a veces atajos) mentales que se emplean para la resolución de los asuntos que llegan a los tribunales. Para este autor, resulta inadecuado «negar el potencial de la AI por el mero hecho de que no sea, ni creo que llegue a ser nunca, un sistema perfecto; o porque pueda producir, aunque sea minimizados, determinados sesgos. Y esto máxime cuando al mismo tiempo no somos tan exigentes con el ser humano, al que admitimos la imperfección como algo natural, y le toleramos decisiones basadas prácticamente en la mera intuición cuando no en el prejuicio más o menos disimulado. En suma, el potencial de la AI requiere ser valorado con criterios y exigencias equivalentes a las que exigimos al ser humano» (21) . Con este reconocimiento de sesgos en el trabajo de los jueces y la preocupación por la discriminación algorítmica, sería conveniente fomentar la formación en psicología dentro del ámbito de la judicatura para prevenir la discriminación judicial derivada del uso de estereotipos que puedan debilitar el derecho a la tutela judicial efectiva.

VII. Soberanía en la era digital

Ya es conocida la expresión de que los datos son el nuevo oro. Lo que pareciera un eslogan es capaz de reconfigurar las dinámicas de poder a nivel nacional y supranacional, replanteando las tensiones en el triángulo de actores formado por el gobierno, el mercado y la ciudadanía. Gobernar esos datos masivos se ha convertido en una prioridad (y/o necesidad), ya que son fuente de riqueza y poder. Si la propiedad de los macrodatos y la capacidad de procesamiento reside en un subgrupo de empresas transnacionales, ¿cuál es la capacidad de influencia de los Estados? ¿Deben los gobiernos reapropiarse de esa información? ¿Se deben crear organismos independientes para la realización de auditorías que comprueben la adecuación de los sistemas de IA al ordenamiento jurídico? ¿Deben los gigantes de Silicon Valley rendir cuentas a la ciudadanía sobre el uso de sus datos?

Siendo imprescindible dar respuesta a estos interrogantes desde un sentido político y con miras a su regulación, de

poco sirve realizar un esfuerzo democrático por colectivizar los datos si no se tiene la capacidad de procesarlos. Utilizando el símil del código abierto, la importancia de hacer accesible un código a la población se desvanece si no se dominan lenguajes de programación. Como también se critica al Derecho, las fórmulas enrevesadas y la terminología compleja en ocasiones aleja a la ciudadanía de cuestiones trascendentales para su vida, siendo la tarea de los juristas traducir las leyes a un lenguaje común. Por tanto, habrá que explicar las matemáticas que hay detrás de los algoritmos y, sin alcanzar un nivel técnico, comprender el funcionamiento de los sistemas de IA para decidir o no cederles la toma de decisiones.

VIII. La sustitución algorítmica en la Administración de Justicia

El paradigma de la sustitución de los seres humanos por las máquinas genera todo tipo de alarmas, desde el reemplazo de la mano de obra y el aumento de la tasa de desempleo, hasta el empobrecimiento de las relaciones al sustituir el factor humano de cualquier interacción, sea para la reclamación de una tarifa telefónica, la compra de un kilo de tomates, el pago de una multa o la prescripción de un medicamento. En el ámbito de la Administración de Justicia este relevo se ve impedido por el principio de exclusividad previsto en los artículos 117.3 de la Constitución Española (LA LEY 2500/1978) y 2.1 de la Ley Orgánica del Poder Judicial (LA LEY 1694/1985) según el cual el ejercicio de la potestad jurisdiccional en todo tipo de procesos, juzgando y haciendo ejecutar lo juzgado, corresponde exclusivamente a los Juzgados y Tribunales determinados por las leyes. Pese a ello, y dado que ya se conocen experiencias en otros países, es posible anticipar algunas de las formas en las que gradualmente se podrían dictar resoluciones «robóticas». BONET NAVARRO realiza una propuesta progresiva considerando que será posible verificar requisitos como la presentación de documentos y copias requeridas en plazo, pero también fijar los hechos controvertidos, valorar la prueba y dictar el fallo de una sentencia fundada en derecho (22) .

No obstante, es preciso realizar dos matizaciones. En primer lugar, respecto a la progresiva algoritmización de la justicia, hay que partir del concepto de algoritmo. Un algoritmo se puede definir como una secuencia de instrucciones que conducen a la resolución de un problema. Ese conjunto de instrucciones ordenadas está presente en un modelo de ML, pero también es frecuente el uso de la algoritmia, en sentido amplio, para dirimir cuestiones de la vida cotidiana. Se emplean algoritmos al diagnosticar un paciente, en una hoja de Excel y en el proceso mental para la toma de decisiones utilizando a veces estructuras selectivas que siguen la siguiente fórmula: Si [A] entonces [B], en caso contrario [C]. Este modo de articular el pensamiento no es ajeno al razonamiento de los jueces: Si [elementos del tipo] entonces [delito] en caso contrario [absolución], o bien, Si [violencia o intimidación] entonces [agresión sexual] en caso contrario [abuso sexual]. La presencia de algoritmos en los tribunales es algo común por lo que el anuncio de la algoritmización de la Justicia tendrá que formularse en relación a la tipología e incidencia del algoritmo (23) , pero no en cuanto al proceso de cumplimiento de órdenes sucesivas que es básico en el pensamiento humano.

En segundo lugar, en caso de producirse el reemplazo de los operadores jurídicos por sistemas de IA se trataría de una sustitución supervisada. No va a llevarse a cabo ninguna transformación sin el control de los sistemas de IA por parte de equipos interdisciplinarios de expertos. Es decir, la progresiva automatización de los tribunales será implantada de acuerdo con las oportunidades y límites que como sociedad concedamos a la IA.

IX. Conclusiones: ¿mito o realidad?

La aproximación a la IA y a sus múltiples aplicaciones, algunas de ellas en el ámbito del derecho, debería realizarse con la máxima rigurosidad, conociendo los fundamentos que vertebran su desarrollo. De otro modo, las críticas a estos sistemas de IA se vuelven peregrinas e infundadas, movidas aún más por un desconocimiento científico-tecnológico o por la idealización de las virtudes asociadas al ser humano que por una oposición genuina. Además, las expectativas sobre estos modelos deben dimensionarse adecuadamente para no exigir criterios inalcanzables. Este artículo pretende ser una contribución al entendimiento de la IA y el Derecho, desmitificando ciertos pensamientos en torno a los sistemas de IA que distorsionan los beneficios y los inconvenientes de su incorporación en el ámbito de la Justicia.

El acercamiento integral y con rigor al tándem IA-Derecho se consigue acotando los conceptos y perfilando los márgenes de los potenciales usos de la IA en los Tribunales. Para ello, de igual modo que se exige el trabajo conjunto de expertos multidisciplinares para la implementación de modelos de ML, es imprescindible contar con un enfoque técnico-jurídico en las investigaciones. De lo contrario, los mitos superarán a la realidad, aproximándonos idealmente a un escenario de ciencia ficción y desperdiciando la contribución que la ciencia y la tecnología pueden

aportar a la sociedad. Lejos de destruir la humanidad, la Inteligencia Artificial nos brinda herramientas para la consecución de un mundo más justo, una aspiración que sin duda nos hace más humanos.

X. Bibliografía

- AGUIRRE-ESPINOSA, Juan Esteban, «Reflexiones sobre el derecho y el juez algoritmo», *Revista Ratio Juris*, vol. 15, núm. 30, 2020, pp. 272-276.
- ANZOVINI, Maria, FERSINI, Elisabetta and ROSSO Paolo, «Automatic Identification and Classification of Misogynistic Language on Twitter», in SILBERZTEIN, Max, ATIGUI, Faten, KORNYSHOVA, Elena, MÉTAIS, Elisabeth and MEZIANE, Farid (Eds.), *Natural Language Processing and Information Systems. 23rd Int. Conf. on Applications of Natural Language to Information Systems*, NLDB-2018, Springer-Verlag, LNCS(10859), pp. 57-64.
- BATELLI, Ettore, «La decisión robótica: algoritmos, interpretación y justicia predictiva», *Revista de Derecho Privado*, núm. 38, 2020, pp. 45-86.
- BONET NAVARRO, José, «¿Puede ser más justa, en materia penal, la inteligencia artificial?», *REVISTA INMEXIUS. Revista de Derecho Penal y Procesal Penal*, núm. 40, 2020, pp. 16-24.
- «La tutela judicial de los derechos no humanos. De la tramitación electrónica al proceso con robots autónomos», *CEFLegal: revista práctica de derecho. Comentarios y casos prácticos*, núm. 208, 2018, pp. 55-92.
- BOSTROM, Nick and YUDKOWSKY, Eliezer, «The ethics of artificial intelligence», in FRANKISH, Keith and RAMSEY, William M. (Eds.), *The Cambridge handbook of artificial intelligence*, Cambridge University Press, 2014, pp. 316-334.
- CHEN, Daniel L., «Machine Learning and the Rule of Law», *Law as Data*, núm. 16, 2019, pp. 1-12.
- COPPIN, Ben, *Artificial Intelligence Illuminated*, Jones and Bartlett Publishers, Massachusetts, 2004.
- CORSO, Anthony, MOSS, Robert J., KOREN, Mark, LEE, Ritchie, KOCHENDERFER, Mykel J., «A Survey of Algorithms for Black-Box Safety Validation of Cyber-Physical Systems», *Journal of Artificial Intelligence Research*, núm. 72, 2021, pp. 377-428. arXiv preprint arXiv:2005.02979
- DE LA OLIVA SANTOS, Andrés (2019): «Justicia predictiva», interpretación matemática de las normas, sentencias robóticas y la vieja historia del «justizklavier», *El Cronista del Estado Social y Democrático de Derecho*, núm. 80, 2019, pp. 30-37.
- DE LA ROSA RODRÍGUEZ, Paola y SANDOVAL NAVARRO, Víctor David, «Los sesgos cognitivos y su influjo en la decisión judicial. Aportes de la psicología jurídica a los procesos penales de corte acusatorio», *Revista Derecho Penal y Criminología*, Vol. 37, núm. 102, 2016, pp. 141-164.
- GOLDENBERG, S. Larry, NIR, Guy and SALCUDEAN, Septimiu E., «A new era: artificial intelligence and machine learning in prostate cancer», *Nature Reviews Urology*, Vol. 16, 2019, pp. 391-403.
- HOIS, Joana, THEOFANOU-FUELBIER, Dimitra and JUNK, Alischa Janine, «How to Achieve Explainability and Transparency in Human AI Interaction» in STEPHANIDIS, Constantine (Ed.) *International Conference on Human-Computer Interaction*, Springer, Cham, 2019, pp. 177-183.
- HUANG, Ming-Hui and RUST, Roland T., «Artificial intelligence in service», *Journal of Service Research*, Vol. 21, núm. 2, 2018, pp. 155-172.
- KARNOUSKOS, Stamatias, «Symbiosis with artificial intelligence via the prism of law, robots, and society», *Artificial Intelligence and Law*, 2021, pp. 1-23.
- MAHESH, Batta, «Machine Learning Algorithms-A Review», *International Journal of Science and Research*, núm. 9, 2020, pp. 381-386.
- MUÑOZ ARANGUREN, Arturo, «La influencia de los sesgos cognitivos en las decisiones jurisdiccionales: el factor humano. Una aproximación», *Indret: Revista para el Análisis del Derecho*, núm. 2, 2011, pp. 1-39.
- SÁNCHEZ-JUNQUERA, Javier, ROSSO, Paolo, MONTES-Y-GÓMEZ, Manuel and CHULVI, Berta, «Masking and BERT-based Models for Stereotype Identification», *Procesamiento del Lenguaje Natural*, núm. 67, 2021, pp. 83-94.
- SORIANO ARNANZ, Alba, «Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos», *Revista de Derecho Público: Teoría y Método*, vol. 3, 2021, pp. 85-127.

SORIANO ARNANZ, Alba, «Decisiones automatizadas y discriminación: aproximación y propuestas generales», *Revista General de Derecho Administrativo*, núm. 56, 2021, pp. 1-45.

SOURDIN, Tania, «Judge v Robot? Artificial Intelligence and judicial decision-making», *UNSW Law Journal*, vol. 41, 2018, pp. 1114-1133.

ULLAH, Zaib, AL-TURJMAN, Fadi, MOSTARDA, Leonardo and GAGLIARDI, Roberto, «Applications of artificial intelligence and machine learning in smart cities», *Computer Communications*, Vol. 154, 2020, pp. 313-323.

VIVES ANTÓN, Tomás Salvador, «La responsabilidad de los jueces en el proyecto de Ley Orgánica del Poder Judicial», *Estudios Penales y Criminológicos*, vol. IX Cursos e Congresos núm. 40. Servizo de Publicacións da Universidade de Santiago de Compostela, 1986, pp. 259-282.

WATSON, David S., KRUTZINNA, Jenny, BRUCE, Ian N., GRIFFITHS, Christopher. E., MCINNES, Ian. B., BARNES, Michael R., and FLORIDI, Luciano, «Clinical applications of machine learning algorithms: beyond the black box» *Bmj*, núm. 364, 2019, pp. 1-4.

WORTHY CAMPBELL, Ray, «Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning», *Colorado Technology Law Journal*, vol. 18, núm. 2, 2020, pp. 323-350.

ZERILLI, John, KNOTT, Alistair, MACLAURIN, James and GAVAGHAN, Colin, «Transparency in algorithmic and human decision-making: is there a double standar?», *Philosophy & Technology*, Vol. 32, núm. 4, pp. 661-683.

(1) GOLDENBERG, S. Larry, NIR, Guy and. SALCUDEAN, Septimiu E., «A new era: artificial intelligence and machine learning in prostate cancer», *Nature Reviews Urology*, Vol. 16, 2019, p. 391.

(2) COPPIN, Ben, *Artificial Intelligence Illuminated*, Jones and Bartlett Publishers, Massachusetts, 2004, p. 5 and HUANG, Ming-Hui and RUST, Roland T., «Artificial intelligence in service», *Journal of Service Research*, Vol. 21, num. 2, 2018, p. 159.

(3) MAHESH, Batta, «Machine Learning Algorithms-A Review», *International Journal of Science and Research*, num.9, 2020, pp. 381-383 and ULLAH, Zaib, AL-TURJMAN, Fadi, MOSTARDA, Leonardo and GAGLIARDI, Roberto, «Applications of artificial intelligence and machine learning in smart cities», *Computer Communications*, Vol. 154, 2020, p. 315.

(4) WORTHY CAMPBELL, Ray (2020): «Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning», *Colorado Technology Law Journal*, vol. 18, num. 2, p. 325.

(5) BOSTROM, Nick and YUDKOWSKY, Eliezer, «The ethics of artificial intelligence», in FRANKISH, Keith and RAMSEY, William M. (Eds.), *The Cambridge handbook of artificial intelligence*, Cambridge University Press, 2014, p. 317.

(6) Para una aproximación al tema consultar SORIANO ARNANZ, Alba, «Decisiones automatizadas y discriminación: aproximación y propuestas generales», *Revista General de Derecho Administrativo*, núm. 56, 2021, pp. 1-45

(7) SORIANO ARNANZ, Alba, «Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos», *Revista de Derecho Público: Teoría y Método*, vol. 3, 2021, p. 92.

(8) VIVES ANTÓN, Tomás Salvador, «La responsabilidad de los jueces en el proyecto de Ley Orgánica del Poder Judicial», *Estudios Penales y Criminológicos*, vol. IX Cursos e Congresos núm. 40. Servizo de Publicacións da Universidade de Santiago de Compostela, 1986, pp. 262-264.

(9) DE LA ROSA RODRÍGUEZ, Paola y SANDOVAL NAVARRO, Víctor David, «Los sesgos cognitivos y su influjo en la decisión judicial. Aportes de la psicología jurídica a los procesos penales de corte acusatorio», *Revista Derecho Penal y Criminología*, Vol. 37, núm. 102, 2016, p. 152.

(10) KARNOUSKOS, Stamatis, «Symbiosis with artificial intelligence via the prism of law, robots, and society», *Artificial Intelligence and Law*, 2021, p. 8, CHEN, Daniel L., «Machine Learning and the Rule of Law», *Law as Data*, num. 16, 2019, pp. 2 and 6 and SOURDIN, Tania, «Judge v Robot? Artificial Intelligence and judicial decision-making», *UNSW Law Journal*, vol. 41, 2018, pp. 1128-1129.

(11) ANZOVINO, Maria, FERSINI, Elisabetta and ROSSO Paolo, «Automatic Identification and Classification of Misogynistic Language on Twitter», in SILBERZTEIN, Max, ATIGUI, Faten, KORNYSHOVA, Elena, MÉTAIS, Elisabeth and MEZIANE, Farid (Eds.), *Natural Language Processing and Information Systems. 23rd Int. Conf. on Applications of Natural Language to Information Systems*, NLDB-2018, Springer-Verlag, LNCS(10859), pp. 57-64.

(12) DE LA OLIVA SANTOS, Andrés (2019): «Justicia predictiva», interpretación matemática de las normas, sentencias robóticas y la vieja historia del «justizklavier», *El Cronista del Estado Social y Democrático de Derecho*, núm. 80, 2019, p. 35.

(13) AGUIRRE-ESPINOSA, Juan Esteban, «Reflexiones sobre el derecho y el juez algoritmo», *Revista Ratio Juris*, vol. 15, núm. 30, 2020, p. 274.

(14) BATELLI, Ettore, «La decisión robótica: algoritmos, interpretación y justicia predictiva», *Revista de Derecho Privado*, núm. 38, 2020, p. 57.

(15) WORTHY CAMPBELL, Ray (2020): «Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning», *op. cit.*, p. 348.

-
- (16) CORSO, Anthony, MOSS, Robert J., KOREN, Mark, LEE, Ritchie, KOCHENDERFER, Mykel J., «A Survey of Algorithms for Black-Box Safety Validation of Cyber-Physical Systems», *Journal of Artificial Intelligence Research*, num. 72, 2021, p. 381. arXiv preprint arXiv:2005.02979.
-
- (17) WATSON, David. S., KRUTZINNA, Jenny, BRUCE, Ian N., GRIFFITHS, Christopher. E., MCINNES, Ian. B., BARNES, Michael R., and FLORIDI, Luciano, «Clinical applications of machine learning algorithms: beyond the black box» *Bmj*, num. 364, 2019, p. 1.
-
- (18) Sobre la identificación de estereotipos hacia migrantes prestando atención a la explicabilidad de los modelos para que los resultados sean comprensibles, resulta pertinente consultar el trabajo de SÁNCHEZ-JUNQUERA, Javier, ROSSO, Paolo, MONTES-Y-GÓMEZ, Manuel and CHULVI, Berta, «Masking and BERT-based Models for Stereotype Identification», *Procesamiento del Lenguaje Natural (SEPLN)*, núm. 67, 2021, pp. 83-94.
-
- (19) HOIS, Joana, THEOFANOU-FUELBIER, Dimitra and JUNK, Alischa Janine, «How to Achieve Explainability and Transparency in Human AI Interaction» in STEPHANIDIS, Constantine (Ed.) *International Conference on Human-Computer Interaction*, Springer, Cham, 2019, p. 182 and ZERILLI, John, KNOTT, Alistair, MACLAURIN, James and GAVAGHAN, Colin, «Transparency in algorithmic and human decision-making: is there a double standard?», *Philosophy & Technology*, Vol. 32, num. 4, pp. 661-683.
-
- (20) MUÑOZ ARANGUREN, Arturo, «La influencia de los sesgos cognitivos en las decisiones jurisdiccionales: el factor humano. Una aproximación», *Indret: Revista para el Análisis del Derecho*, núm. 2, 2011, p. 16.
-
- (21) BONET NAVARRO, José, «¿Puede ser más justa, en materia penal, la inteligencia artificial?», *REVISTA INMEXIUS. Revista de Derecho Penal y Procesal Penal*, núm. 40, 2020, p. 20.
-
- (22) BONET NAVARRO, José, «La tutela judicial de los derechos no humanos. De la tramitación electrónica al proceso con robots autónomos», *CEFLegal: revista práctica de derecho. Comentarios y casos prácticos*, núm. 208, 2018, pp. 80 y ss.
-
- (23) Si bien la IA está compuesta por algoritmos, no todos los algoritmos son IA. De igual forma, mientras que los algoritmos clásicos necesitan de antemano instrucciones prefijadas, los modelos de ML buscan por sí solos los patrones para la obtención de un resultado.
-