

RETOS PARA UNA INTELIGENCIA ARTIFICIAL INCLUSIVA
DE LOS COLECTIVOS VULNERABLES

*CHALLENGES FOR AN INCLUSIVE ARTIFICIAL INTELLIGENCE OF
VULNERABLE GROUPS*

Actualidad Jurídica Iberoamericana N° 21, agosto 2024, ISSN: 2386-4567, pp. 360-383

Ana Isabel
BLANCO
GARCÍA

ARTÍCULO RECIBIDO: 29 de mayo de 2023

ARTÍCULO APROBADO: 1 de julio de 2024

RESUMEN: La revolución tecnológica de la inteligencia artificial es una realidad innegable. Todos los sistemas, aplicaciones, herramientas... que calificamos como inteligentes están llamados a transformar los actuales modelos de convivencia, de comunicación e, incluso, de tutela de nuestros derechos. Ahora bien, a pesar del progreso y avance que ello está suponiendo en nuestra sociedad, como sucede con todo cambio trascendental, son muchos los desafíos que genera, especialmente si queremos garantizar un modelo de Justicia caracterizado por la inclusión y la igualdad.

PALABRAS CLAVE: Artificial Intelligence; algorithms; vulnerable group; equality; inclusion.

ABSTRACT: *Technological Revolution of Artificial Intelligence is an undeniable reality. All systems, applications, tools... qualified as intelligent are meant to transform the current models of living and communication, and even the protection of human rights. However, despite the progress and advances in our society, there are many challenges as a transcendental change, especially when the aspiration is an inclusive and fair model of justice.*

KEY WORDS: *Inteligencia Artificial; algoritmos; colectivos vulnerables; igualdad; inclusión.*

SUMARIO.- I. PUNTO DE PARTIDA: HACIA UNA INTELIGENCIA ARTIFICIAL INCLUSIVA, CONFIABLE Y ÉTICA.- II. LA IA Y LOS ODS: SU NECESARIA ALINEACIÓN.- 1. Primer reto: sesgos algorítmicos.- 2. Segundo reto: trazabilidad y explicabilidad del algoritmo.- 3. Tercer reto: la motivación de las resoluciones.- III. LA IA Y LA PERSPECTIVA DE VULNERABILIDAD: UNA REFLEXIÓN FINAL.

I. PUNTO DE PARTIDA: HACIA UNA INTELIGENCIA ARTIFICIAL INCLUSIVA, CONFIABLE Y ÉTICA.

La revolución tecnológica ha sufrido, en los últimos tiempos, un salto cualitativo gracias a la irrupción -masiva, atropellada y algo caótica- de la Inteligencia Artificial (en adelante, IA), donde su presencia es ya una constante en prácticamente todos los ámbitos y sectores de la sociedad. Atrás vamos dejando las confusiones sobre qué es la IA y sobre lo que no es IA para adentrarnos de lleno en su necesaria -y esperada- regulación que permita sentar las bases de un marco jurídico eficaz. Así es, la novedad de los sistemas informáticos y de la digitalización de los procesos ha quedado superada. Y, obviamente, el ámbito de la Justicia no ha quedado al margen de esta modernización, donde ya se aplican sistemas y herramientas de IA.

La utilización de sistemas de IA por parte de entidades de sectores productivos y económicos de forma transversal a lo largo de la cadena de valor conlleva una serie de consideraciones de relevancia para el ámbito jurídico. La mayoría cumplen finalidades como la prevención del fraude, del blanqueo de capitales o de comisión de delitos dentro de la propia organización. Otras se configuran como *softwares* asistenciales -los conocidos como Chatbots- e incluso también pueden servir para resolver quejas y reclamaciones sencillas y que permitan una estandarización en la reparación de los daños causados a las personas usuarias de tales servicios o productos.

Indudable es, también, el potencial de las herramientas de IA en la mejora, agilización y sostenibilidad del sistema de Justicia actual, gracias no solamente a la propia aplicación de tecnología sino también a la automatización de los procedimientos y distintas actuaciones, lo que permitirá una redistribución más eficiente de los recursos disponibles.

Nos hallamos, pues, en un escenario en el que el progreso y transformación del paradigma de Justicia se muestra favorable a la calidad, seguridad y eficiencia de los procesos y medios de tutela. Sin embargo, la otra cara de la moneda, menos amable, viene representada por todos los riesgos inherentes a la aplicación de estas tecnologías y su posible afectación de derechos fundamentales.

• **Ana Isabel Blanco García**

Profesora Titular de Derecho procesal, Universitat de València. Correo electrónico: a.isabel.blanco@uv.es

Una serie de desafíos que derivan de la implantación de esta IA y de las funciones variadas que desempeña, desde la mejora del acceso a la justicia, la asistencia jurídica, la resolución de conflictos a través de plataformas online¹, hasta la función jurisdiccional, donde puede cumplir funciones predictivas y funciones decisorias en el seno del proceso judicial², auxiliando en distintos ámbitos como la valoración del riesgo, la valoración de la prueba o la calificación jurídica³.

Con carácter previo a la identificación y estudio de estos desafíos, es importante entender qué es una herramienta de IA, para así poder mostrar cómo podemos tratar de preservar la equidad y la igualdad de la sociedad, de los valores, de los principios y de los derechos de las personas.

Concretamente, se entiende que la IA engloba “sistemas con la capacidad de realizar funciones asociadas a la inteligencia humana como percibir, aprender, entender, adaptarse, razonar e interactuar imitando un comportamiento humano inteligente”⁴. Por su parte, y más recientemente, el Reglamento de Inteligencia Artificial, aprobado por el Parlamento Europeo el día 13 de marzo de 2024⁵ viene a armonizar el concepto de IA⁶, pues hasta la fecha se contaba con distintas definiciones. Así, este texto concibe la IA como “un sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras el despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar

- 1 En palabras de BARONA, el desarrollo de una plataforma digital ODR es una fórmula que “rompe el sistema de resolución de conflictos e introduce la algoritmización en la toma de decisiones al trabajar con la plataforma digital no solo en lo formal sino también en lo material, en la solución del conflicto en sí”. BARONA VILAR, S.: “Inteligencia artificial o la algoritmización de la vida y de la justicia”, *Revista Boliviana de Derecho*, núm. 28, 2019, p. 40. Sobre la aplicación de la IA en los ODR, véase, MARCOS FRANCISCO, D.: “Smart ODR y su puesta en práctica: el salto a la inteligencia artificial”, *Revista General de Derecho Procesal*, 2023, núm. 59.
- 2 MARTÍN DÍZ, F.: “Justicia digital post-covid19: el desafío de las soluciones extrajudiciales electrónicas de litigios y la inteligencia artificial”, *Revista de Estudios Jurídicos y Criminológicos*, 2020, núm. 2, p. 63. Sobre la repercusión jurídica de las herramientas de IA en las distintas fases del proceso, véase, MONTESINOS GARCÍA, A.: “Empleo de la inteligencia artificial en algunas fases del proceso judicial civil: prueba, medidas cautelares y sentencia”, *Actualidad civil*, 2022, núm. 11, pp. 1-31.
- 3 CASTILLEJO MANZANARES, R.: “Las nuevas tecnologías y la inteligencia artificial como retos post-covid19”, *Revista General de Derecho Procesal*, 2022, núm. 56, p. 12.
- 4 OLIVER, N., *Inteligencia Artificial, naturalmente. Un manual de convivencia entre humanos y máquinas para que la tecnología nos beneficie a todos*, ONTSI, Madrid, 2020, p. 28. Disponible en: <https://www.unav.edu/documents/26661763/0/InteligenciaArtificialNuriaOliver.pdf/b527df6-b04a-025b-2000-8ba3d157d1ae?t=1625572030005>
- 5 European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). Documento disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html#title2
- 6 BARONA VILAR la concibe como “una suerte de noción omnicomprendensiva (panconcepto) que permite abrigar desde las primeras manifestaciones de la máquina inteligente, siquiera lo fuere para desplegar una suerte de talento en una determinada ciencia, arte, cultura, hasta cuestionarnos, en este imparable mundo expansivo digital que vivimos, nuevas fórmulas inteligentes, no solo en cuanto a tecnología de última generación, sino también en cuanto a su accesibilidad y amabilidad de uso”. BARONA VILAR, S.: *Algoritmización del Derecho y de la Justicia. De la Inteligencia Artificial a la Smart Justice*, Tirant lo Blanch, Valencia, 2021, p. 80.

información de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos o virtuales” (artículo 3.1).

Esta tecnología inteligente diluye la separación y límites entre la tecnología y las personas, entre el factor tecnológico y el factor humano. Vivimos inmersos en el fenómeno de “la algoritmización de la vida”⁷, que nos avizora para cambiar nuestro entorno, pero también la forma de crear, interpretar y aplicar el Derecho. Tanto es así que ya se habla incluso de la aparición de un nuevo sujeto o figura como es el juez robot o juez IA, definido por GÓMEZ COLOMER como “una máquina inteligente que dicta sentencias con base en los algoritmos introducidos en ella y en los hechos y datos del caso real producido”⁸, de forma que puedan automatizarse procedimientos sencillos⁹, dado que han demostrado resultar útiles, ágiles y eficaces. No obstante, no podemos olvidar que deberá, en todo caso, delimitarse la relación entre el factor tecnológico y el factor humano dada la prohibición de que la IA ocupe “el lugar de un ser humano a la hora de dictar sentencia o tomar decisiones”¹⁰.

Sobre este aspecto, en el marco de un proceso penal, BARONA insiste en la necesidad de valorar si el juez robot podría llegar a sustituir al juez humano en el proceso penal. Al respecto, considera que “extrapolar estas iniciativas de robotización judicial a la sede penal llevaría a la liquidación del proceso penal, al convertir este en un expediente automatizado que impediría el ejercicio de los derechos reconocidos a quienes son sujetos del proceso penal; mermarían las garantías de defensa y propulsarían una mecanización judicial que, cuando menos, chocaría con el proceso lógico jurídico de razonamiento judicial que aplica la norma al caso concreto, al sujeto concreto y bajo unas circunstancias concretas, modulando y justificando esta modulación en la motivación de los hechos probados”¹¹.

En suma, la incorporación de estos sistemas reporta beneficios en términos de simplicidad, agilización y reducción de costos. Ahora bien, como toda novedad,

7 BARONA VILAR, S.: “Inteligencia artificial”, cit., p. 23.

8 GÓMEZ COLOMER, J. L.: “Unas reflexiones sobre el llamado «juez-robot», al hilo del principio de independencia judicial”, en AA.VV.: *Justicia algorítmica y neodercho: una mirada multidisciplinar* (dir. por S. BARONA VILAR), Tirant lo Blanch, Valencia, 2021, p. 258.

9 CONDE FUENTES se muestra crítico con la automatización de los procesos judiciales al cuestionarse qué debería entenderse por sencillo y el alcance de la automatización, dudando de su efectividad en la fase decisoria. CONDE FUENTES, J.: “La inteligencia artificial y la figura del juez-robot”, en AA.VV.: *Modernización, eficiencia y aceleración del proceso* (dir. por S. PEREIRA PUIGVERT y M.J. PESQUEIRA ZAMORA), Aranzadi, Cizur Menor (Navarra), 2022, pp. 118-119.

10 Resolución del Parlamento Europeo, de 20 de enero de 2021, sobre inteligencia artificial: cuestiones de interpretación y de aplicación del Derecho internacional en la medida en que la UE se ve afectada en los ámbitos de los usos civil y militar, así como de la autoridad del Estado fuera del ámbito de la justicia penal, ap. 69. (2020/2013(INI)). (DOUE C 456/04, de 10 de noviembre de 2021).

11 BARONA VILAR, S.: “Cuarta revolución industrial (4.0.) o ciberindustria en el proceso penal: revolución digital, inteligencia artificial y el camino hacia la robotización de la justicia”, *Revista Jurídica Digital UANDES*, 2019, vol. 3, núm. 1, pp. 15-16.

también presenta múltiples desafíos en relación con su propia naturaleza y con el impacto de sus aplicaciones, especialmente en lo que concierne a la transparencia, la privacidad de los datos, la gobernanza y el entrenamiento de los algorítmicos. Lo que más preocupa es su posible contribución a la amplificación de los sesgos presentes en los datos que los alimentan, exacerbando así las desigualdades, las brechas sociales y la inequidad. Por ello reviste de gran importancia la necesidad de asentar los cimientos y andamiaje de estos sistemas inteligentes par poder desarrollar un entorno confiable.

Precisamente, la confiabilidad de la IA ha sido una preocupación de organismos internacionales, como la Organización para la Cooperación y el Desarrollo Económicos quien, mediante la adopción, en mayo de 2019, de los Principios de la OCDE sobre la Inteligencia Artificial¹², busca proteger los estándares internacionales dirigidos a la construcción y despliegue de sistemas de IA fiables, seguros y justos. Los estándares son siempre necesarios, pero aquí devienen esenciales para la aplicación en procesos en los que se hallen vinculadas personas en riesgo de vulnerabilidad o vulnerables, respetando en todo momento la dignidad humana y contribuyendo a una sociedad inclusiva.

En concreto, el Principio 1 ha reconocido que la inteligencia artificial debe fundamentar su desarrollo en el crecimiento inclusivo, el desarrollo sostenible y en el bienestar que, aunado con el Principio 2, se deberá realizar con base en unos valores centrados en el ser humano y la Justicia, donde destaca la importancia de los valores democráticos, la biodiversidad y la equidad para lograr una sociedad más justa. En este progreso no puede olvidarse la necesidad de transparencia y capacidad de explicación de los algoritmos (Principio 3), así como su trazabilidad para poder ponderar los riesgos (Principio 4) y evitar, así, una vulneración o merma de los derechos de las personas y que los sistemas de IA se caractericen por la robustez, la seguridad y la protección. Para su consecución, el Principio 5 cierra el ciclo al exigir la observancia del correcto funcionamiento de la IA al amparo de un régimen de responsabilidad.

Estos Principios deben complementarse con otros dos documentos clave. Por un lado, las Directrices Éticas para una IA fiable, establecidas por la Comisión Europea¹³, que buscan orientar la aplicabilidad de estos sistemas inteligentes. Y, por otro lado, con la Recomendación sobre la ética de la inteligencia artificial de la UNESCO, adoptada el 23 de noviembre de 2021¹⁴, dirigida a garantizar que todas estas transformaciones se alineen con los derechos humanos y los ODS.

12 Documento disponible en: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

13 Documento publicado el 8 de abril de 2019 y elaborado por el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, creado en junio de 2018. Contenido del texto disponible en: <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-81f-01aa75ed71a1>

14 Documento disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Respecto de las Directrices Éticas, es importante señalar que el Grupo Independiente de Expertos resalta la importancia del seguimiento y cumplimiento de tales principios y estándares, en tanto en cuanto “la IA tiene el potencial de transformar significativamente la sociedad. La IA no es un fin en sí mismo, sino un medio prometedor para favorecer la prosperidad humana y, de ese modo, mejorar el bienestar individual y social y el bien común, además de traer consigo progreso e innovación.

En particular, los sistemas de IA pueden ayudar a facilitar el logro de los Objetivos de Desarrollo Sostenible de las Naciones Unidas, como la promoción del equilibrio entre mujeres y hombres y la lucha contra el cambio climático, la racionalización del uso que los seres humanos hacemos de los recursos naturales, la mejora de la salud, la movilidad y los procesos de producción y el seguimiento de los avances en los indicadores de sostenibilidad y cohesión social”¹⁵. Ahora bien, para alcanzar tal fin se torna imprescindible desarrollar protocolos y mecanismos para controlar y garantizar la fiabilidad de la IA, minimizando los potenciales riesgos que podrían colocar en una situación altamente perjudicial a las personas vulnerables o potencialmente vulnerables.

En este sentido, una IA fiable debe fundarse en estándares y valores que garanticen el respeto de los derechos fundamentales mediante la prevención de los riesgos inherentes a la utilización de este tipo de herramientas inteligentes, no solo a través de la protección de los principios de igualdad y equidad de las personas, sino también de neutralidad tecnológica, evitándose cualquier tipo de discriminación arbitraria, siendo imprescindible una mayor exigencia de transparencia y trazabilidad de la herramienta, sin olvidar el irrenunciable respeto de la autonomía humana en la adopción de las decisiones.

En la misma aparece la “Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno”¹⁶, identificando una serie de principios que pretender actuar como guía de la implementación de la IA en el ámbito judicial, a saber: respeto por los derechos fundamentales; prohibición de discriminación entre personas o grupos de personas; calidad y seguridad en los datos que sirvan para configurar los algoritmos; la transparencia, imparcialidad y justicia de los códigos y lenguaje del algoritmo y, finalmente, el control de su correcto funcionamiento.

Con todo ello vuelve a quedar demostrada la relación y necesaria colaboración entre los factores tecnológico y humano para evitar que la IA se convierta en un

15 Directrices Éticas para una IA fiable, ap. 9.

16 Adoptada por el Grupo de Trabajo sobre la Calidad de la Justicia (CEPEJ-GT-QUAL) del Consejo de Europa en diciembre de 2018.

instrumento discriminatorio¹⁷. Para lograr este objetivo de establecer un marco para una IA fiable se deben garantizar principios como el respeto de los derechos humanos en la implementación sus servicios, tales como la protección de datos y de privacidad, la diversidad, la no discriminación y la equidad, al mismo tiempo que se asegure una gestión eficiente de la calidad de los sistemas, donde prime la seguridad y la imparcialidad de la decisión algorítmica.

Por lo tanto, es cierto que el uso responsable y ético de la IA contribuye a la mejora y rendimiento de procesos automatizados, así como agiliza y simplifica tareas arduas y costosas. Sin embargo, un uso irresponsable de la IA puede tener un efecto negativo y servir para reforzar y consolidar discriminaciones sistémicas, agrandar las brechas sociales, obstaculizar e incluso impedir el acceso de determinadas poblaciones y colectivos a los derechos sociales.

II. LA IA Y LOS ODS: SU NECESARIA ALINEACIÓN.

La Agenda 2030 y sus 17 Objetivos de Desarrollo sostenible¹⁸ (en adelante, ODS) nacieron con la clara voluntad de transformar la sociedad, su pensamiento, su filosofía, su forma de hacer y entender la realidad..., pero siempre con la mira puesta en un objetivo claro: la sostenibilidad económica, social y ambiental. Con esta finalidad ha diseñado una hoja de ruta que viene delimitada por los 17 ODS y 169 metas que “hace necesario transformar el paradigma de desarrollo dominante en uno que nos lleve por la vía del desarrollo sostenible, inclusivo y con visión de largo plazo” y donde se “pone a la igualdad y dignidad de las personas en el centro”¹⁹.

En este compromiso universal, llamado a mejorar nuestra sociedad, la IA puede jugar, y así lo hará, un papel determinante en la consecución y logro de las distintas metas proyectadas. Efectivamente, la tecnología más avanzada puede ponerse al servicio de las Agencias y Organizaciones Internacionales, de Administraciones, de ONGs, etc. y contribuir a la optimización de todos los análisis e indicadores, así como mejorar los procesos y tiempos de respuesta a los problemas y desafíos identificados.

En el ámbito de la Justicia son varias las muestras de lo que la IA puede contribuir en la mejora y agilización de los procesos. No obstante, aunque vamos poco a poco

17 Para MONTESINOS, la responsabilidad de la decisión final “debe recaer siempre en una persona”. MONTESINOS GARCÍA, A.: “Inteligencia Artificial y ODR”; en AA.VV.: *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*” (dir. por S. BARONA VILAR), Tirant lo Blanch, Valencia, 2021, p. 527.

18 El contenido de cada uno de los ODS establecidos en la Agenda 2030 se puede consultar en el siguiente enlace: <https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/>

19 Agenda 2030, pág. 7. Documento disponible en: <https://repositorio.cepal.org/server/api/core/bitstreams/cb30a4de-7d87-4e79-8e7a-ad5279038718/content>

conociendo el poder, la capacidad, el alcance y el impacto de las herramientas de IA desarrolladas, su increíble potencial hace que continúe siendo desconocida en muchos aspectos. Igualmente, como todo avance y novedad, habida cuenta de su gran proyección en cuanto a su colaboración con los operadores jurídicos, son varios y complejos los desafíos que presenta.

En este trabajo hemos querido resaltar aquellos retos que, de no afrontarlos, profundizarían y perpetuarían las brechas y desigualdades sociales, perjudicando sobremanera a personas tradicionalmente castigadas por la discriminación y prejuicios de las sociedades. Estos desafíos pivotan en torno a la configuración de la IA mediante algoritmos, esas operaciones matemáticas complejas y tan poco comprensibles. En este sentido, la manera en que los algoritmos se nutren de datos y cómo los gestiona y analiza puede no ser la mejor ni la más adecuada al hallarse sesgos o elementos discriminatorios.

La posible existencia de sesgos algorítmicos se une a la opacidad de su creación y configuración, lo que dificulta la explicabilidad del algoritmo y, con ello, la posible quiebra de principios esenciales como el de motivación de las resoluciones si se deja la toma de decisión completamente en manos de la máquina. El equilibrio entre el factor humano y el factor tecnológico es clave para el correcto funcionamiento de estas herramientas y sistemas de IA.

I. Primer reto: sesgos algorítmicos.

El beneficio que puede aportar una herramienta de IA se verá diluido si su configuración perpetúa, consolida y agrava las desigualdades de nuestra sociedad. Los algoritmos se nutren de "un conjunto finito de reglas/comandos, generalmente en la forma de una lógica matemática, que permite obtener un resultado a partir de elementos de entrada"²⁰. Sin embargo, todos sabemos que "ni la neutralidad maquínica es tan neutral, ni la neutralidad humana lo es"²¹. Las personas tenemos sesgos, prejuicios, opiniones, valores... que en no pocas ocasiones definen nuestra conducta, criterios o juicios, de forma que debemos trabajar en evitar impregnar de tal subjetividad a la máquina.

A tal fin, disponer de un buen *dataset* deviene crucial para el desarrollo de esta tecnología. Ahora bien, la compilación de información que servirá para alimentar la herramienta plantea una doble dificultad. Por un lado, la protección del derecho a la privacidad de los datos e información, muchas veces sensible. Para evitar la manipulación indebida o la comercialización de tales datos, deberá regularse la

20 Council of Europe Commissioner for Human Rights: *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, 2019, p. 24. Documento disponible en: <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>

21 BARONA VILAR, S.: *Algoritmización del Derecho*, cit., p. 618.

recogida de datos buscando el equilibrio entre el derecho a la protección de datos y privacidad y la necesidad de configurar debidamente la herramienta de IA. Por otro lado, la codificación del lenguaje “inteligente” a partir de información de la realidad. Deberá atenderse a una correcta gestión y tratamiento de datos con el fin de contar con información completa, suficiente y carente de vicios o prejuicios, pues de lo contrario los resultados que arroje la máquina serán erróneos.

La codificación del lenguaje se antoja compleja dado que no todas las circunstancias pueden reducirse a números ni tampoco objetivarse hasta el punto de que la estandarización consecuente sea, al final, una mera generalización de los supuestos, sin atender a las particularidades de los distintos casos. Asimismo, la propia naturaleza de la IA permite su aprendizaje a partir de los datos introducidos o *feedback* y la interacción con el entorno. Este riesgo de sesgos o desviaciones se acentúa cuando se trata de herramientas de *Deep learning*, siendo actualmente la mayor preocupación la eliminación de los posibles sesgos.

Estas ecuaciones matemáticas analizan un importante número de datos²², entre los que encontramos información sobre género, edad, nacionalidad e incluso, en ocasiones, creencias religiosas o ideológicas, extraídas de evidencias empíricas existentes, para poder arrojar un resultado -o una predicción- acerca de la probabilidad de que concurra una situación de riesgo, pero al hacerlo pueden servirse de criterios y soluciones pasadas que podían contener sesgos que ahora estas máquinas replicarán, excepto si se aplica un criterio o factor corrector para evitar esa codificación del pasado²³.

La existencia de sesgos es una de las grandes preocupaciones no solo de la utilización de la IA, sino también de su regulación. Inquietud compartida por la Comisión y expresamente recogida en las Directrices “Generar confianza en la inteligencia artificial centrada en el ser humano”²⁴, en concreto en lo que concierne precisamente a la “Diversidad, no discriminación y equidad” (título V).

La Comisión manifiesta que “[L]os conjuntos de datos utilizados por los sistemas de IA (tanto para el entrenamiento como para el funcionamiento) pueden verse afectados por la inclusión de sesgos históricos involuntarios, por no estar completos o por modelos de gobernanza deficientes. La persistencia en estos sesgos podría dar lugar a una discriminación (in)directa. También pueden producirse daños por

22 PÉREZ ESTRADA, J.: “La inteligencia artificial como prueba científica en el proceso penal español”, *Revista Brasileira de Direito Processual Penal*, 2021, vol. 7, núm. 2, p. 1392.

23 Manifiestan la misma preocupación por la mirada de la técnica al pasado, BATELLI, E.: “La decisión robótica: algoritmos, interpretación y justicia predictiva”, *Revista de Derecho Privado*, 2020, núm. 38, p. 51 y NIEVA FENOLL, J.: “Inteligencia Artificial y Proceso Judicial: perspectivas tras un alto tecnológico en el camino”, *Revista General de Derecho Procesal*, 2022, núm. 57, p. 16.

24 Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: *Generar confianza en la inteligencia artificial centrada en el ser humano* (COM/2019/168 final).

la explotación intencionada de sesgos (del consumidor) o por una competencia desleal. Por otra parte, la forma en la que se desarrollan los sistemas de IA (por ejemplo, la forma en que está escrito el código de programación de un algoritmo) también puede estar sesgada. Estos problemas deben abordarse desde el inicio del desarrollo del sistema”.

Una propuesta de solución para resolver estos problemas de sesgos algorítmicos sería el establecimiento de equipos de diseño heterogéneos y diversificados²⁵, que atendieran a la realidad plural y la diversidad sexual, racial, cultural que existe en el mundo. Igualmente, resulta de lo más conveniente y oportuno consultar a las partes interesadas que puedan verse afectadas por el sistema. Los sistemas de IA deberían atender a todo el abanico de capacidades, aptitudes, habilidades y necesidades humanas y aplicar un enfoque universal que permitiera alcanzar la igualdad de acceso a todas las personas.

Comprobamos, pues, que el sesgo algorítmico puede tener varias causas, que LAZCOZ²⁶ agrupa en tres categorías y con gran impacto en la exacerbación de las desigualdades.

En primer lugar, “sesgos introducidos por datos incorrectos, irrelevantes o incompletos”, que aluden a la insuficiencia del *dataset*, tanto a nivel cualitativo como cuantitativo. Por un lado, la base de datos puede no ser lo suficientemente completa como para constituir una muestra significativa porque un determinado colectivo haya quedado infrarrepresentado o “porque, en ocasiones, esos datos no constituyan reflejo fiel del espectro total de situaciones y realidades que se pretenden aprehender, al centrarse sólo en alguna o algunas de ellas”²⁷. Por otro lado, aun cuando se disponga de un amplio conjunto de datos, puede suceder que no reflejen todo el abanico de situaciones y circunstancias (de vulnerabilidad), lo que ofrecería una visión parcial de la realidad (del vulnerable), no pudiendo analizar correctamente todas las situaciones.

En segundo lugar, “sesgos por una distribución desigual real de las variables” o “sesgos sociales”, que son aquellos datos que reflejan situaciones sociales históricamente desiguales o discriminatorias y que son replicados al establecerse perfiles o patrones de conducta, lo que perpetuaría la situación discriminatoria, especialmente con personas de raza negra, migrantes, mujeres, o personas con discapacidad.

25 Con el mismo parecer, SCASSERA, S.: “La desigualdad automatizada. Industrialización, exclusión y colonialismo digital”, *Nueva Sociedad*, 2021, núm. 294, 2021, p. 51.

26 LAZCOZ MORATINOS, G.: “Modelos algorítmicos, sesgos y discriminación”, en AA.VV.: *FODERTICS 9.0. Estudios sobre tecnologías disruptivas y justicia* (dir. por F. BUENO DE MATA), Comares, Granada, 2021, pp. 286-287.

27 SAIZ GARITAONANDIA, A.: “Personas vulnerables, justicia e inteligencia artificial. Motivos para permanecer alerta”, cit., p. 99.

Finalmente, en tercer lugar, reconoce los “sesgos relacionados con la transformación de los datos para facilitar su procesamiento posterior por el modelo algorítmico”, esto es, que el tratamiento y gestión de los datos que van a introducirse o han sido introducidos para su posterior análisis automatizado no sea el idóneo.

En consecuencia, la replicación o creación de sesgos y prejuicios por una mala configuración del *dataset* o del propio algoritmo preocupa sobremedida por su afectación y consecuencias negativas a los colectivos tradicionalmente oprimidos y vulnerables, lo que provocaría un recrudecimiento de la violencia, abuso y desigualdad estructural. Se trata de evitar, en definitiva, que los estereotipos ya existentes en nuestra sociedad puedan trasladarse, y perpetuarse, en el mundo digital.

Llama la atención el impresionante desarrollo de sistemas de IA en el orden penal, configurándose para atender distintos fines, tales como la predicción del riesgo de reincidencia, de delitos e incluso de fraudes -financieros o a la Seguridad Social-, entre otras funciones. Ahora bien, el ejemplo paradigmático de la existencia de sesgos en estas herramientas lo constituye el caso *Wisconsin vs. Loomis*²⁸, donde se demostró la existencia de un sesgo por razón del origen racial de las personas en la herramienta COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), en la que un algoritmo predictivo de aprendizaje automático o *machine learning* adoptado por la Administración de Justicia de Estados Unidos arrojaba porcentajes de riesgo de reincidencia mucho mayores cuando la persona sospechosa era de raza negra.

Este suceso refleja no solo el problema de la opacidad de la configuración del algoritmo, sino también la asunción de la discriminación histórica de las personas de raza negra, sin la debida justificación de la necesidad de tomar en consideración la raza para el análisis y sin la combinación de dicho factor con otros que resultaren oportunos para verificar el nivel de riesgo como acusado. Llama la atención el contraste entre el refuerzo de los derechos y garantías de la persona acusada y la utilización de algoritmos predictivos que mantienen y reproducen las desigualdades estructurales de la sociedad. Unas distinciones (arbitrarias) que buscamos erradicar precisamente con la (supuesta) objetivación de los aspectos a tomar en consideración para estimar un determinado nivel de riesgo o la culpabilidad.

28 *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

Para un análisis de esta resolución judicial y sus implicaciones en el proceso, véase, por todos, GARCÍA SÁNCHEZ, M.D.: “Retos del uso de la Inteligencia Artificial en el proceso: impugnaciones con fundamentación algorítmica y Derecho a la tutela judicial efectiva”, en AA.VV.: *FODERTICS 9.0. Estudios sobre tecnologías disruptivas y Justicia* (dir. por F. BUENO DE MATA), Comares, Granada, 2021, pp. 233-244.

Por otra parte, encontramos también ejemplos de discriminaciones en el uso de la IA en otros ámbitos, no judiciales, pero con repercusiones jurídicas. En concreto en el ámbito del crédito, donde los sesgos son el resultado del análisis automatizado, por ejemplo, del nivel de riesgo para la concesión del crédito. A tal fin, las entidades elaboran un perfil²⁹ de la persona solicitante para lo que utilizan criterios tales como el nivel de renta, el historial crediticio o su capacidad de pago, pero también se ha descubierto que se tienen en cuenta factores como la edad o el género³⁰.

Desde instancias supranacionales se viene instando a todas las organizaciones a erradicar este tipo de discriminaciones injustificadas, y más aún evitar su reproducción en las herramientas de IA. Sin embargo, como podemos comprobar, las desigualdades y desequilibrios continúan, pues factores como el género no se entiende que sigan teniéndose en cuenta porque no aportan ningún tipo de factor ni elemento que pudiera afectar a la resolución de concesión o no de un crédito o de cualquier otro tipo de financiación.

Al contrario, simplemente sirve para perpetuar estereotipos que aumentan la brecha de género existente, especialmente en personas de mayor edad. Ello hace que, por ejemplo, personas con menores recursos económicos o que hubieran sido incluidas anteriormente en listados de morosidad sean inmediatamente excluidas o vean sus condiciones menos ventajosas, lo que agravaría su situación financiera e incrementaría el nivel de riesgo de exclusión social.

No olvidemos tampoco que la salud ha sido un factor considerado para la concesión de créditos, negándose automáticamente a personas con enfermedades como el cáncer, aun habiéndolo superado. Dos han sido las importantes consecuencias de esta realidad discriminatoria por culpa de la algoritmización.

La primera, el nacimiento de un nuevo derecho, el llamado “derecho al olvido oncológico”³¹, que refiere a la potestad conferida a las personas a solicitar la

29 De conformidad con el artículo 4.4) del 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), debemos entender por elaboración de perfiles “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”. (DOUE L 119/1, de 4 de mayo de 2016).

30 La propia Recomendación sobre la ética de la inteligencia artificial, adoptada el 23 de noviembre de 2021 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura ya recomendaba que “Los Estados Miembros deberían velar por que los estereotipos de género y los sesgos discriminatorios no se trasladen a los sistemas de IA, sino que se detecten y corrijan de manera proactiva” (ap. 90). Documento disponible en el siguiente enlace: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

31 BARONA señala que, junto a este nuevo derecho, son múltiples las repercusiones jurídicas que tiene esta revolución digital, tales como “la regulación de las cookies y el marketing relacional, el régimen legal de las aplicaciones, la economía colaborativa, el ciberespacio, la cibercriminalidad, la ciberseguridad, la

supresión de información personal y datos relacionados, tal y como recoge el Reglamento General de Protección de Datos (RGPD). En concreto, podrán pedirlo pacientes que hayan superado un cáncer, siempre y cuando hubieran transcurrido 5 años desde la finalización del tratamiento sin recaída en el momento de la suscripción del crédito, seguro o hipoteca. Estamos ante la eliminación de una desigualdad histórica, cuyos efectos se han extendido también a otro colectivo tradicionalmente marginado y reprimido como las personas con la enfermedad VIH/SIDA.

Precisamente, desde instancias supranacionales se ha abordado esta causa de discriminación que tantos perjuicios estaba ocasionando para el desarrollo de las personas con normalidad, dictando la Resolución del Parlamento Europeo, de 16 de febrero de 2022, sobre el refuerzo de Europa en la lucha contra el cáncer: hacia una estrategia global y coordinada³², aprobado el 16 de febrero de 2022, donde se indica que “Considerando que el objetivo del Plan no debe ser solo luchar contra un problema de salud pública esencial y ayudar a los pacientes a vivir más y mejor, sino que también debe ser iniciar una reducción de las desigualdades e injusticias y reducir la carga social y económica de la enfermedad; que la Comisión debe promover un enfoque centrado en el paciente y basado en los derechos de los ciudadanos integrando consideraciones de justicia, sostenibilidad, equidad, solidaridad, innovación y colaboración en el núcleo mismo del Plan, incluida su Iniciativa para Ayudar a los Niños con Cáncer” (ap. E).

Para dar cumplimiento con esta Resolución, que instaba a reforzar la lucha contra el cáncer mediante una estrategia global y coordinada en toda Europa, el 27 de junio de 2023 se aprobó por el Consejo de Ministros el Real Decreto-ley 5/2023, de 28 de junio, por el que se adoptan y prorrogan determinadas medidas de respuesta a las consecuencias económicas y sociales de la Guerra de Ucrania, de apoyo a la reconstrucción de la isla de La Palma y a otras situaciones de vulnerabilidad; de transposición de Directivas de la Unión Europea en materia de modificaciones estructurales de sociedades mercantiles y conciliación de la vida familiar y la vida profesional de los progenitores y los cuidadores; y de ejecución y cumplimiento del Derecho de la Unión Europea³³, que puso en marcha este derecho al olvido oncológico.

protección de los derechos fundamentales en internet, la *eJustice*, las nuevas técnicas de investigación tecnológicamente avanzadas, la incidencia del *Big Data* en todas las áreas jurídicas, la regulación de internet y especial protección de los menores en internet, el internet de las cosas, entre otros tantos. BARONA VILAR, S.: *Algoritmización del Derecho*, cit., p. 73.

32 2020/2267(INI).

33 BOE núm. 154, de 29 de junio de 2023.

El Preámbulo, apartado III, señala que “el capítulo II incorpora medidas para hacer efectivo el derecho al olvido en la contratación de seguros y productos bancarios de los pacientes de patologías oncológicas una vez transcurrido un determinado período de tiempo desde la finalización del tratamiento sin recaída. Para ello se establece, por un lado, la nulidad de las cláusulas que excluyan a una de las partes por haber padecido cáncer; y, por otra, la prohibición de discriminación en la contratación de un seguro a una persona

En concreto, destacan los artículos 209, apartado Dos, por el que se modifica la DA 5ª de la Ley 50/1980, de 8 de octubre, de Contrato de Seguro y artículo 210, por el que se modifica la DA Única del Texto Refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias, aprobado por Real Decreto Legislativo 1/2007, de 16 de noviembre³⁴.

De conformidad con esta modificación, ya “no se podrá discriminar a las personas que tengan VIH/SIDA, ni por otras condiciones de salud. En particular, se prohíbe la denegación de acceso a la contratación, el establecimiento de procedimientos de contratación diferentes de los habitualmente utilizados por el asegurador o la imposición de condiciones más onerosas, por razón de tener VIH/SIDA, o por otras condiciones de salud, salvo que se encuentren fundadas en causas justificadas, proporcionadas y razonables, que se hallen documentadas previa y objetivamente”.

Igualmente señala la nulidad, en relación con las cláusulas no negociadas individualmente, de “aquellas cláusulas, estipulaciones, condiciones o pactos que excluyan a una de las partes” bien por tener VIH/SIDA u otras condiciones de salud, bien por ser supervivientes de un cáncer antes de la fecha de suscripción del contrato o negocio jurídico, introduciendo un factor temporal de cinco años desde la finalización del tratamiento radical sin recaída posterior.

Esto es importante para la configuración de la herramienta de IA correspondiente, que “no puede estar diseñada, ni ser aplicada, generando discriminaciones, incurriendo en la creación de perfiles con sesgo, marginación o exclusión de colectivos especialmente vulnerables. El principio de igualdad, la presunción de inocencia, la protección de la privacidad y de los datos personales, así como el pleno ejercicio del derecho de defensa y el derecho a un proceso justo que no pueden verse afectados en ningún caso. Para ello es determinante la observancia de los Códigos éticos y sus principios desde las estructuras institucionales o privadas que apliquen sistemas de inteligencia artificial en el ámbito de la justicia”³⁵.

Dada la utilidad y funcionalidad de la IA, SAIZ propone “sugerir el uso de la propia IA de cara a poder detectar esos sesgos. Su gran capacidad de proceso de datos y determinación de patrones puede ser de gran ayuda para percibir

por haber sufrido una patología oncológica, una vez transcurridos, en ambos casos, cinco años desde la finalización del tratamiento radical sin recaída posterior. Además, para suscribir un seguro de vida tampoco habrá obligación de declarar si se ha padecido cáncer una vez cumplido el mencionado plazo, ni se podrán tomar en consideración dichos antecedentes oncológicos, a estos efectos”.

34 BOE núm. 287, de 30 de noviembre de 2007.

35 MARTÍN DIZ, F.: “Capítulo 45. Inteligencia artificial y proceso: garantías frente a eficiencia en el entorno de los derechos procesales fundamentales”, en AA.VV.: *Justicia: ¿garantías “versus” eficiencia?* (coord. por R. BELLIDO PENADÉS), Tirant lo Blanch, Valencia, 2019, p. 822.

los *bias* ocultos en las decisiones pasadas, ser conscientes de los mismos y ponerlos remedio”³⁶. Ello sería posible en tanto en cuanto las conclusiones o resultados que arrojan los algoritmos son resultado de complejas ecuaciones y operaciones que tratan de hallar patrones de conducta, establecer modelos de respuesta y, al hacerlo, pueden también estar replicando los sesgos introducidos por las personas, pudiendo incluso amplificarlos si no somos capaces de detectar el error y la desigualdad al manejar y analizar grandes cantidades de datos. De momento, empero, no contamos con algoritmos cuya finalidad sea la de eliminar las desigualdades existentes en género, raza, etnia, estatus socioeconómico, entre otros muchos factores.

La segunda de las consecuencias de este cambio normativo refiere del planteamiento sobre quién asumiría la responsabilidad de la vulneración del derecho a no ser objeto de discriminación arbitraria por motivos de raza, género, etnia, nivel económico, nivel educativo, o enfermedad. La atribución de los efectos jurídicos de la decisión o acción automatizada que origina una discriminación a ciertos colectivos no es baladí.

RODRÍGUEZ DE LAS HERAS apunta a que tal responsabilidad se atribuiría al operador “en la medida en que controla (o debería poder controlar) los riesgos de operar un sistema de IA que decide integrar en su actividad y, por tanto, beneficiarse de su funcionamiento”³⁷ siendo, en este caso, la propia entidad la infractora del principio de no discriminación.

La falta de exhaustividad del marco jurídico de la IA, aún en construcción, y el exigido respeto a la dignidad, la igualdad y la no discriminación de las personas, sienta la necesidad de articular cómo “deberán diseñarse, implementarse y ponerse en servicio sistemas de IA para satisfacer los principios de trazabilidad, explicabilidad, transparencia, supervisión humana, auditabilidad, no discriminación, explicación razonada de las decisiones, o acceso a un mecanismo de revisión de decisiones significativas”³⁸, y a quiénes deberá atribuirse la responsabilidad (civil, penal, administrativa...) -y, correlativamente, la legitimación pasiva- en caso de infracción y discriminación.

Consecuencia de todo ello, podemos confirmar que la revolución que la IA ha supuesto -no solo en el ámbito de la Justicia sino también en general- y el elevado

36 SAIZ GARITONANDIA, A.: “Personas vulnerables, justicia e inteligencia artificial. Motivos para permanecer alerta”, en AA.VV.: *Personas vulnerables y tutela penal* (dir. por N.J. DE LA MATA BARRANCO y A.I. PÉREZ MACHÍO), Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2023, p. 102. En la misma línea se pronuncia FERRANTE, E.: “Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos?”, *Nueva Sociedad*, 2021, núm. 294, p. 34.

37 RODRÍGUEZ DE LAS HERAS BALLELL, T.: “Inteligencia Artificial en el sector bancario: reflexiones sobre su régimen jurídico en la Unión Europea”, *ICE. Revista de Economía*, 2022, núm. 926, p. 103.

38 *Ibidem*, p. 105.

grado de fiabilidad que muchos sistemas han demostrado, servirá para alcanzar los ODS de la Agenda 2030, diseñando un sistema más justo e inclusivo para la ciudadanía. No obstante, su introducción en todos los ámbitos de la sociedad, pero en especial de la Justicia, debe realizarse de una forma garantista, tuitiva de los Derechos Fundamentales. A tal fin, reiteramos que la prevalencia del factor humano en el control y toma de decisiones deviene clave para lograr una justicia inclusiva, segura y confiable, precisamente para asegurar la protección y defensa, en términos de igualdad, de personas vulnerables o en una situación de vulnerabilidad.

2. Segundo reto: trazabilidad y explicabilidad del algoritmo.

La IA aporta un gran avance y cuenta con una importante proyección, pero también tiene limitaciones, entre las que destaca la opacidad del contenido del algoritmo. Precisamente, esta opacidad o falta de transparencia se equipara al concepto de caja negra³⁹ porque resulta muy complejo comprender y conocer los entresijos del algoritmo.

En otras palabras, la dificultad de la utilización de algoritmos o sistemas predictivos reside en la necesidad de mayor transparencia (ante la completa falta de ella) y, correlativamente, del conocimiento de la lógica del algoritmo y de la capacidad de justificación de la decisión. Por una parte, la transparencia en el tratamiento de datos, en la configuración y en la técnica empleada por el algoritmo debe ser un principio de actuación de la Administración de Justicia, sin olvidarnos, por otra parte, de la búsqueda de un equilibrio entre este principio y el derecho de protección de datos.

La administración de justicia ante una eventual vulneración de derechos por la existencia de sesgos en la IA, por ejemplo, se dificulta sobremanera dada la poca confiabilidad y transparencia del algoritmo. Estas exigencias de transparencia y fiabilidad obedecen a la necesidad de disponer de una explicación de la decisión que permita conocer el proceso de análisis y entender así el resultado que haya emitido la herramienta inteligente.

Es más, la falta de cualquiera de estos requisitos podría suponer la ilegalidad del propio algoritmo, tal como ha sucedido con SyRI, dedicado a la evaluación de características personales de los ciudadanos, declarado ilegal por el Tribunal de Distrito de la Haya, en su Sentencia de 5 de febrero de 2020⁴⁰, anteriormente

39 PRICE, W.N.: "Artificial Intelligence in Health Care: Applications and Legal Issues", *U of Michigan Public Law Research Paper*, 2017, núm. 599, p. 2.

40 Asunto C/09/550982/HA ZA 18-388 (NJCM et al./Países Bajos. Sobre esta Resolución, véase, COTINO HUESO, L.: "SyRI, ¿a quién sanciono? Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020", *La Ley Privacidad*, 2020, núm. 4.

utilizado por parte del Gobierno de los Países Bajos para combatir el fraude a la Seguridad Social. En consecuencia, la trazabilidad y explicabilidad del algoritmo permite lograr el objetivo de disponer de una IA confiable, dado que la confianza se basa en conocer cómo funciona el algoritmo y cuáles son los parámetros que considerar cuando analiza un supuesto concreto.

El Libro Blanco de la Unión Europea sobre Inteligencia Artificial aborda esta opacidad y la problemática que subyace para los derechos humanos, alegando que “[L]as características particulares de numerosas tecnologías de IA, como la opacidad («efecto caja negra»), la complejidad, la imprevisibilidad y un comportamiento parcialmente autónomo, pueden hacer difícil comprobar el cumplimiento de la legislación vigente de la UE sobre la protección de los derechos fundamentales e impedir su cumplimiento efectivo. Puede ser que las fuerzas y cuerpos de seguridad y las personas afectadas carezcan de los medios para comprobar cómo se ha tomado una decisión determinada con ayuda de la IA y, por consiguiente, si se han respetado las normas pertinentes. Las personas físicas y las personas jurídicas pueden enfrentarse a dificultades en el acceso efectivo a la justicia en situaciones en las que estas decisiones les afecten negativamente”⁴¹.

Preocupación por la “transparencia e inteligibilidad del funcionamiento de los algoritmos y los datos con los que han sido entrenados” que también recoge la UNESCO debido a su impacto “en, entre otros, la dignidad humana, los derechos humanos y las libertades fundamentales, la igualdad de género, la democracia, los procesos sociales, económicos, políticos y culturales, las prácticas científicas y de ingeniería, el bienestar animal y el medio ambiente y los ecosistemas”⁴².

La solución pasaría por disponer de *softwares* abiertos que permitan conocer los *inputs* e información analizada, auditar los datos y poder cuestionar el resultado. Solo así se garantizarían los derechos y se evitaría la manipulación, mala interpretación o error en los *outputs*.

3. Tercer reto: la motivación de las resoluciones.

Mucho se habla de la automatización de los procesos, de la funcionalidad de las herramientas de IA, e incluso de la figura del Juez robot, esto es, del papel que debe desempeñar la IA en el sistema de justicia. Cuestión nada baladí, porque aunque parezca una realidad lejana, está cada día más próxima.

41 Libro Blanco de la Comisión Europea sobre Inteligencia Artificial titulado “Una aproximación europea a la excelencia y a la confianza”, de 19 de febrero de 2020. COM (2020) 65 final, pp. 14-15.

42 UNESCO, *Recomendación sobre la ética de la inteligencia artificial*, 2021, p. 1. [SHS/BIO/REC-AIETHICS/2021]. Documento disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa

Como acertadamente señala MARTÍN DIZ, la IA debe desempeñar, de momento, una labor asistencial⁴³. Las características y funcionamiento de las herramientas de IA justificarían esta necesidad de supervisión por parte del justiciable. En este sentido, como afirma GUZMÁN FLUJA, el “desarrollo de la IA general y el continuo avance y perfeccionamiento en la programación de algoritmos, así como la depuración en cantidad y calidad de los datos a manejar, y la mejora del autoaprendizaje de las IA”⁴⁴ requieren del factor humano, lo que impediría, en un escenario futuro, prescindir del justiciable para la toma de la decisión.

Con todo, tampoco podemos despistarnos y olvidarnos de que la exigencia de motivación de las decisiones adoptadas en el marco de un proceso es un derecho comprendido en el derecho de acceso a la tutela judicial efectiva (artículo 24 de la Constitución), que establece la obligación del justiciable de ofrecer una “respuesta razonada, motivada y congruente con las pretensiones oportunamente deducidas por las partes”, así como una exigencia consagrada en el artículo 120.3 y susceptible, en consecuencia, de amparo constitucional.

Una obligación que no desaparece con la utilización de una IA, más bien al contrario, debe reforzarse, ya que los *outputs* de la IA pueden “contribuir a motivar más y mejores decisiones ahora difíciles, ambivalentes, que en ocasiones se adoptan en el vacío o sin suficiente información fruto de la incapacidad de tener en cuenta ciertas variables o datos relacionados”⁴⁵.

En este sentido, conocer cuál ha sido el razonamiento y la interpretación que se ha dado de las circunstancias para determinar la concurrencia de uno o varios riesgos se convierte en el eje para garantizar el derecho a un proceso justo y el derecho de defensa. Por el contrario, desconocer cómo y por qué se ha emitido una probabilidad, un resultado o una decisión que afecta a los derechos de las personas, entrañaría un grave riesgo para la seguridad jurídica y para los principios procesales.

Aquí es donde entendemos, y vemos muy claro, que el factor humano no puede -o no debería- ser reemplazado por una máquina quien, de momento, no es capaz de replicar la función argumentativa del órgano juzgador, por lo que

43 MARTÍN DIZ, F., “Modelos de aplicación de Inteligencia Artificial en justicia: asistencial o predictiva versus decisoria”, en AA.VV.: *Justicia algorítmica y neuroderecho: una mirada multidisciplinar* (ed. por S. BARONA VILAR), Tirant lo Blanch, Valencia, 2021, pp. 65-85.

44 GUZMÁN FLUJA, V. C., “Proceso penal y justicia automatizada”, *Revista General de Derecho Procesal*, 2021, núm. 53, p. 33.

45 SIMÓN CASTELLANO, P.: *Justicia cautelar e Inteligencia Artificial*, Bosch, Madrid, 2021, p. 71. En el mismo sentido se pronuncia PÉREZ RAGONE, A.: “Justicia civil en la era digital y artificial: ¿Hacia una nueva identidad?”, *Revista Chilena de Derecho*, 2021, vol. 48, núm. 2, p. 215. y PLANCHADELL GARGALLO, A.: “Inteligencia Artificial y medidas cautelares”, en AA.VV.: *Justicia algorítmica y neuroderecho, Una mirada multidisciplinar* (ed. por S. BARONA VILAR.), Tirant Lo Blanch, Valencia, 2021, p. 410.

hablaríamos de un escenario donde el juzgador y la máquina de IA pudieran convivir para una mejora del sistema de justicia.

III. LA IA Y LA PERSPECTIVA DE VULNERABILIDAD: UNA REFLEXIÓN FINAL.

La introducción de aplicaciones de IA presenta un gran potencial y una impresionante proyección por su utilidad y eficacia para el progreso, pero también por su versatilidad, por su capacidad de adaptación y contribución al progreso en sectores como la medicina, la agricultura, la ciencia, la educación, etc., lo que nos permite afirmar que será una gran contribución y ayuda en la consecución de los ODS.

Sin embargo, todo este profundo impacto positivo en las sociedades y en su desarrollo económico, político, social y ambiental, quedara ensombrecido -y eclipsado en algunos aspectos- por el agravamiento de las desigualdades sociales ya existentes, reforzando la marginalización de algunas poblaciones.

El temor a los sistemas de IA no proviene tanto de la posibilidad de que las máquinas inteligentes reemplacen a las personas habida cuenta de la mayor optimización y eficacia de los procesos -algo que lleva sucediendo décadas- sino por el hecho de que estos sistemas tratan de emular y, en cierta manera, de reproducir, el pensamiento humano. Ahora bien, la gran diferencia que impide que esa aspiración se torne realidad, es la imaginación, la creatividad y la capacidad de crítica del ser humano y de las que la máquina aún carece. De ahí que la máquina se limite a valorar un resultado mediante la estandarización y patronaje de similares escenarios.

Aquí surge el mayor desafío. Si nos preguntamos si la IA puede llegar a agravar, exacerbar o generar situaciones de desequilibrio y desigualdad, deberíamos responder afirmativamente -al menos, de momento-. La privacidad de datos, la accesibilidad a la tecnología, la falencia de competencias digitales, entre otros, son elementos a considerar para evitar agravar las brechas sociales, la exclusión social y la discriminación, esto es, para impedir que colectivos que son actualmente vulnerables lo sigan siendo por culpa del uso de una IA que pretende ser útil, fiable y eficaz pero que, en realidad, no respeta los valores éticos ni la igualdad entre las personas.

Debe, por tanto, llevarse a cabo una reflexión profunda, sosegada, y bajo una perspectiva de vulnerabilidad, sobre la configuración y utilización de la IA en el ámbito judicial, estableciéndose límites inquebrantables que defiendan la legalidad, la igualdad y la no discriminación entre las personas.

BIBLIOGRAFÍA

BARONA VILAR, S.: "Cuarta revolución industrial (4.0.) o ciberindustria en el proceso penal: revolución digital, inteligencia artificial y el camino hacia la robotización de la justicia", *Revista Jurídica Digital UANDES*, 2019, vol. 3, núm. 1, pp. 1-21.

BARONA VILAR, S.: "Inteligencia artificial o la algoritmización de la vida y de la justicia", *Revista Boliviana de Derecho*, 2019, núm. 28.

BARONA VILAR, S.: *Algoritmización del Derecho y de la Justicia. De la Inteligencia Artificial a la Smart Justice*, Tirant lo Blanch, Valencia, 2021.

BATELLI, E.: "La decisión robótica: algoritmos, interpretación y justicia predictiva", *Revista de Derecho Privado*, 2020, núm. 38.

CASTILLEJO MANZANARES, R.: "Las nuevas tecnologías y la inteligencia artificial como retos post-covid19", *Revista General de Derecho Procesal*, 2022, núm. 56.

CONDE FUENTES, J.: "La inteligencia artificial y la figura del juez-robot", en AA.VV.: *Modernización, eficiencia y aceleración del proceso* (dir. por S. PEREIRA PUIGVERT y M.J. PESQUEIRA ZAMORA), Aranzadi, Cizur Menor (Navarra), 2022, pp. 115-136.

COTINO HUESO, L.: "SyRI, ¿a quién sanciono? Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020", *La Ley Privacidad*, 2020, núm. 4.

FERRANTE, E.: "Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos?", *Nueva Sociedad*, 2021, núm. 294, pp. 27-36.

GARCÍA SÁNCHEZ, M.D.: "Retos del uso de la Inteligencia Artificial en el proceso: impugnaciones con fundamentación algorítmica y Derecho a la tutela judicial efectiva", en AA.VV.: *FODERTICS 9.0. Estudios sobre tecnologías disruptivas y Justicia* (dir. por F. BUENO DE MATA), Comares, Granada, 2021, pp. 233-244.

GÓMEZ COLOMER, J. L.: "Unas reflexiones sobre el llamado «juez-robot», al hilo del principio de independencia judicial", en AA.VV.: *Justicia algorítmica y neuroderecho: una mirada multidisciplinar* dir. por S. BARONA VILAR), Tirant lo Blanch, Valencia, 2021, pp. 243-263.

GUZMÁN FLUJA, V. C.: "Proceso penal y justicia automatizada", *Revista General de Derecho Procesal*, 2021, núm. 53.

LAZCOZ MORATINOS, G.: "Modelos algorítmicos, sesgos y discriminación", en AA.VV.: *FODERTICS 9.0. Estudios sobre tecnologías disruptivas y Justicia* (dir. por F. BUENO DE MATA), Comares, Granada, 2021, p. 283-294.

MARCOS FRANCISCO, D.: "Smart ODR y su puesta en práctica: el salto a la inteligencia artificial", *Revista General de Derecho Procesal*, 2023, núm. 59.

MARTÍN DIZ, F.: "Capítulo 45. Inteligencia artificial y proceso: garantías frente a eficiencia en el entorno de los derechos procesales fundamentales", en AA.VV.: *Justicia: ¿garantías "versus" eficiencia?* (coord. por R. BELLIDO PENADÉS), Tirant lo Blanch, Valencia, 2019, pp. 815-827.

MARTÍN DIZ, F.: "Justicia digital post-covid19: el desafío de las soluciones extrajudiciales electrónicas de litigios y la inteligencia artificial", *Revista de Estudios Jurídicos y Criminológicos*, 2020, núm. 2, pp. 41-74.

MARTÍN DIZ, F., "Modelos de aplicación de Inteligencia Artificial en justicia: asistencial o predictiva versus decisoria", en AA.VV.: *Justicia algorítmica y neuroderecho: una mirada multidisciplinar* (ed. por S. BARONA VILAR), Tirant lo Blanch, Valencia, 2021, pp. 65-85.

MONTESINOS GARCÍA, A.: "Empleo de la inteligencia artificial en algunas fases del proceso judicial civil: prueba, medidas cautelares y sentencia", *Actualidad civil*, 2022, núm. 11, pp. 1-31.

MONTESINOS GARCÍA, A.: "Inteligencia Artificial y ODR; en AA.VV.: *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*" (dir. por S. BARONA VILAR), Tirant lo Blanch, Valencia, 2021, pp. 507-531.

NIEVA FENOLL, J.: "Inteligencia Artificial y Proceso Judicial: perspectivas tras un alto tecnológico en el camino", *Revista General de Derecho Procesal*, 2022, núm. 57.

OLIVER, N., *Inteligencia Artificial, naturalmente. Un manual de convivencia entre humanos y máquinas para que la tecnología nos beneficie a todos*, ONTSI, Madrid, 2020.

PÉREZ ESTRADA, J.: "La inteligencia artificial como prueba científica en el proceso penal español", *Revista Brasileira de Direito Processual Penal*, 2021, vol. 7, núm. 2, pp. 1385-1410.

PÉREZ RAGONE, A.: "Justicia civil en la era digital y artificial: ¿Hacia una nueva identidad?", *Revista Chilena de Derecho*, 2021, vol. 48, núm. 2, pp. 203-229.

PLANCHADELL GARGALLO, A.: "Inteligencia Artificial y medidas cautelares", en AA.VV.: *Justicia algorítmica y neuroderecho, Una mirada multidisciplinar* (ed. por S. BARONA VILAR.), Tirant Lo Blanch, Valencia, 2021, pp. 384-419.

PRICE, W.N.: "Artificial Intelligence in Health Care: Applications and Legal Issues", *U of Michigan Public Law Research Paper*, 2017, núm. 599, pp. 1-7.

RODRÍGUEZ DE LAS HERAS BALLELL, T.: "Inteligencia Artificial en el sector bancario: reflexiones sobre su régimen jurídico en la Unión Europea", *ICE. Revista de Economía*, 2022, núm. 926, pp. 93-107.

SAIZ GARITAONANDIA, A.: "Personas vulnerables, justicia e inteligencia artificial. Motivos para permanecer alerta", en AA.VV.: *Personas vulnerables y tutela penal* (dir. por N.J. DE LA MATA BARRANCO y A.I. PÉREZ MACHÍO), Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2023, pp. 93-110.

SCASSERA, S.: "La desigualdad automatizada. Industrialización, exclusión y colonialismo digital", *Nueva Sociedad*, 2021, núm. 294, pp. 49-60.

SIMÓN CASTELLANO, P.: *Justicia cautelar e Inteligencia Artificial*, Bosch, Madrid, 2021.

UNESCO, *Recomendación sobre la ética de la inteligencia artificial*, 2021. [SHS/BIO/REC-AIETHICS/2021].

