

BRECHAS DIGITALES Y SESGOS EN LA INTELIGENCIA ARTIFICIAL

¿Una nueva forma de discriminación?*

DIGITAL DIVIDES AND BIASES IN ARTIFICIAL INTELLIGENCE A new form of discrimination?

M.^a Olga Sánchez Martínez**

RESUMEN: El potencial de la inteligencia artificial, para mejorar determinados aspectos de la vida de las personas, no debe ser un impedimento para adoptar una posición crítica respecto de algunos procedimientos y aplicaciones que puedan afectar negativamente a la igualdad, acentuar discriminaciones preexistentes o propiciar la aparición de nuevas formas de discriminación. Algunas características atribuidas a la inteligencia artificial predisponen a abogar por su desarrollo en un ámbito amplio de libertad, con controles éticos, más que jurídicos, y de *soft law*, en lugar de *hard law*. Sin embargo, cuando se trata de garantizar derechos, como la igualdad, es necesario adoptar medidas jurídicas contundentes, tanto de carácter preventivo como reactivo, frente a las brechas digitales, los sesgos y las discriminaciones que puedan generar.

ABSTRACT: *The potential of artificial intelligence, to improve certain aspects of people's lives, should not be an impediment to taking a critical position on some procedures and applications that adversely affect equality, neither accentuate pre-existing discrimination, nor encourage the emergence of new forms of discrimination. Some characteristics attributed to artificial intelligence predispose us to advocate its development in a wide area of freedom, with ethical, rather than legal, and soft law controls, instead of hard law. However, when it comes to guaranteeing rights such as equality, it is necessary to adopt strong legal measures, both preventive and reactive, to address the digital divides, biases, and discrimination they may generate.*

PALABRAS CLAVE: inteligencia artificial, brechas digitales, sesgos, discriminación algorítmica, accesibilidad.

KEYWORDS: *artificial intelligence, digital divides, biases, algorithmic discrimination, accessibility.*

Fecha de recepción: 5/04/2025

Fecha de aceptación: 13/06/2025

doi: <https://doi.org/10.20318/universitas.2025.9573>

* Esta publicación es parte del proyecto de I+D+i "Inteligencia artificial jurídica y Estado de Derecho" [PID2022 - 139773OB-I00], financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

** Profesora Titular de Filosofía del Derecho. Universidad de Cantabria. E-mail: maria.sanchez@unican.es. ORCID: <https://orcid.org/0000-0001-7641-3125>

1.- INTRODUCCIÓN

En el ámbito de los derechos humanos la universalidad y la igualdad han estado siempre reivindicadas, reconocidas y desarrolladas sobre exclusiones de personas y grupos. La existencia de colectivos que, legítimamente, pretenden ocupar un lugar entre los iguales es una constante histórica. Como constante ha de ser la atención dirigida a cada nueva circunstancia que pueda afectar al ejercicio de los derechos, que son de todas las personas, pero no todas los disfrutan por igual. Porque mientras los derechos se expanden a nuevos sujetos y contenidos, tratando de dar respuesta a las necesidades y exigencias que surgen en los nuevos contextos en que se ejercen, las dificultades sobrevenidas tienden a centrarse en los grupos más vulnerables, en quienes históricamente han sido olvidados y excluidos, los peor tratados, quienes han padecido discriminaciones.

La tecnología digital, sus diversos recursos y aplicaciones, se encuentran entre los elementos que configuran las sociedades del siglo XXI y que son determinantes para el ejercicio de los derechos. En términos de igualdad, las herramientas que componen el universo digital pueden ser utilizadas a su favor, pero también pueden representar una amenaza capaz de generar nuevas formas de discriminación, codificar y ampliar las existentes. Las redes sociales, los procesos de inteligencia artificial, la obtención, el uso y abuso de los datos personales de que se nutre, y los algoritmos que la materializan, pueden incidir en los prejuicios y trasladar resultados perjudiciales hacia determinados colectivos, incluso crear algunos nuevos producto del desarrollo de la misma tecnología, dando lugar a una nueva forma de discriminación: la discriminación algorítmica.

Las diversas funciones de la inteligencia artificial, y su utilización en multitud de tareas, repercuten en la vida de las personas y el ejercicio de sus derechos. En este trabajo, partiendo de los abundantes y variados usos de la inteligencia artificial, se podrán apreciar sus ventajas, pero también los prejuicios que pueden estar presentes en su diseño y ejecución y, como consecuencia, los perjuicios para determinados colectivos. Se pondrán de manifiesto ciertas inexactitudes respecto de algunas características, que se le presuponen a los sistemas de inteligencia artificial, cuyo efecto es limitar la adopción de una necesaria perspectiva crítica respecto a la misma. Los procesos de inteligencia artificial y sus relaciones con la diversidad, las brechas digitales y los sesgos darán cuenta de cómo la gestión de los datos y las fórmulas algorítmicas están en constante tensión con el principio de igualdad y el derecho a la no discriminación. Finalmente, se plantearán algunas de las peculiaridades y dificultades específicas para detectar y corregir la discriminación algorítmica. Tenerlas en cuenta será preciso para establecer medidas eficaces, desde un punto de vista ético, y sobre todo jurídico, a fin de poder garantizar la

igualdad de oportunidades y caminar en la buena dirección para poder disfrutar de los beneficios de la tecnología digital sin discriminaciones.

2.- LA INTELIGENCIA ARTIFICIAL Y SUS MÚLTIPLES USOS

Las nuevas tecnologías y la inteligencia artificial han irrumpido con fuerza en nuestras vidas, afectando con ello al ejercicio de diversos derechos fundamentales. Desde las relaciones más personales hasta el ámbito político, pasando por la sanidad, la educación, el trabajo, las transacciones comerciales, el crédito o el ocio, derechos como la intimidad, el libre desarrollo de la personalidad, la libertad de expresión, el derecho de información, el derecho a la integridad y a la salud, a la educación, al trabajo o la cultura se verán concernidos por la digitalización de la tecnología.

La inteligencia artificial trata de emular a la inteligencia humana para realizar determinadas actividades, algunas mecánicas, otras más complejas, llegando a alcanzar un cierto grado de autonomía. Consiste en un sistema informático que, nutrido de datos y de una secuencia de instrucciones precisas sobre los mismos, es capaz de analizar el entorno, establecer correlaciones, identificar patrones de comportamiento, realizar determinadas tareas, resolver ciertos problemas, hacer predicciones y tomar decisiones. Dos elementos fundamentales componen el motor para poner en marcha los procesos de inteligencia artificial: los datos y los algoritmos. Los datos, proporcionados por las personas al interactuar con la tecnología, son su energía, los nutrientes sobre los que actúan los algoritmos que, siguiendo las instrucciones de los técnicos, los analiza, valora y clasifica, según patrones determinados para la realización de actividades muy diversas, proponer o tomar decisiones.

Los sistemas de inteligencia artificial pueden ser utilizados, entre otros, para fines de asesoramiento, de resolución de conflictos, preparación y emisión de documentos, para conducir automóviles, seleccionar o recomendar determinadas actividades, operar en el sistema financiero, como sistema de identificación, para el reconocimiento de personas, para geolocalizar o seleccionar e indicar rutas en los desplazamientos. Puede expandir la comunicación en diversos formatos, transformar texto oral en lenguaje de signos, la imagen visual y el habla en texto escrito, el escrito en habla, en dibujos o pictogramas, elaborar documentos de lectura fácil, traducir a cualquier idioma, realizar subtítulo automatizado, reconocer la voz e interactuar dando respuesta a peticiones concretas y procurar asistencia para multitud de actividades. Permite adaptarse a las circunstancias particulares y hacer un uso más eficiente de recursos en temas como la salud, la selección de tratamientos, la detección temprana de enfermedades o la prevención. La educación puede

facilitar el proceso de aprendizaje de forma permanente, flexible y adaptada.

Se ha mostrado también como un instrumento muy valioso para detectar necesidades y en el reparto de ayudas sociales. Por eso, el desarrollo digital ha encontrado su acomodo en el Estado del bienestar, que ha visto en la innovación tecnológica un medio para mejorar la eficiencia en la distribución de sus recursos y un instrumento apto para detectar el fraude. Pero esta realidad entraña algunos riesgos asociados a la desinstitucionalización de ciertas funciones antes públicas y que, debido al avance de la inteligencia artificial, se están desplazando hacia entidades privadas. El potencial desarrollo de la inteligencia artificial no está en manos de las administraciones públicas sino de empresas tecnológicas, entre cuyas prioridades no se encuentran los derechos humanos sino la rentabilidad económica, que es la lógica del mercado.

Siendo así, todo proceso de automatización a través de la inteligencia artificial puede tener consecuencias en la actividad cotidiana y en las expectativas de futuro; en las oportunidades que puede brindar y de las que puede privar. La distribución desigual de estas oportunidades, como resultado de los procesos algorítmicos, pueden concluir en la toma de decisiones que colocan, o inciden, en posiciones de desventaja y obstaculizan el ejercicio de derechos de algunos colectivos.

Las posibilidades que estos procedimientos nos ofrecen son múltiples y sus ventajas incontestables, pero también pueden serlo los perjuicios que un uso indebido puede ocasionar. Como portadores y procesadores de datos, para ser utilizados para diversos fines, los algoritmos son transmisores de valor y la inteligencia artificial combina, por tanto, elementos técnicos, sociales y éticos¹. Cathy O'Neil describió los algoritmos como armas de destrucción matemática, capaces de dar lugar a una «producción masiva, prácticamente industrial, de injusticia»²; Yuval Harari advierte que los procesos de inteligencia artificial podrían convertirse en «armas de destrucción masiva sociales», si son capaces de socavar nuestros vínculos de convivencia³. Al respecto, no se debe olvidar que, en una sociedad democrática, el pluralismo, la igualdad, el respeto a la diversidad y la protección de grupos desfavorecidos forman parte de aquellos vínculos de convivencia.

¹ Antonio Luis Terrones Rodríguez, «Inteligencia artificial fiable y vulnerabilidad: una mirada ética sobre los sesgos algorítmicos», en *Vulnerabilidad digital. Desafíos y amenazas de la sociedad hiperconectada*, ed. por Rebeca Suárez Álvarez, Miguel Ángel Martín Córdoba y Luis Manuel Fernández Martínez (Madrid: Dykinson, 2023), 266.

² Cathy O'Neil, *Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y amenaza la democracia*. Trad. por Violeta Arranz de la Torre (Madrid: Capitán Swing, 2017), 120.

³ Yuval Noah Harari, *Nexus. Una breve historia de la información desde la Edad de Piedra hasta la IA*. Trad. por Joanménec Ros, (Barcelona: Debate, 2024), 420.

3.- LOS LÍMITES DE LA INTELIGENCIA ARTIFICIAL Y SUS TENSAS RELACIONES CON LA DIVERSIDAD

Es frecuente que se resalten y reiteren algunas características de la innovación tecnológica digital, más en concreto de la inteligencia artificial, que no se corresponden totalmente con la realidad y pueden ser un lastre para un análisis crítico y riguroso sobre sus perjuicios y el establecimiento de mecanismos de corrección. Entre sus pretendidas, e interesadas, peculiaridades se destacan el carácter inevitable, la neutralidad, imparcialidad u objetividad y la infalibilidad; de semejantes virtualidades se derivará la escasez de mecanismos de autocorrección. Entre los elementos propios, que ahondan en minimizar los aspectos críticos, la complejidad para descifrar sus códigos fuente, y la pobre relación de los sistemas de inteligencia artificial con la diversidad, producirán un inevitable impacto sobre la igualdad, más bien en términos de desigualdad.

En relación con la inevitabilidad, se señala que no existe una alternativa mejor para realizar las funciones y actividades para las que existen herramientas tecnológicas que la que estas ofrecen. Sin embargo, optar por los diversos usos y aplicaciones de la tecnología, o elegir otras opciones, es una decisión humana para la que se tendrían que valorar ventajas e inconvenientes. Por tanto, la tecnología es evitable si se considera que así sea y se prefiere hacer uso de otros recursos disponibles. Otra cosa es que sea útil, eficiente y, especialmente, rentable. En este sentido, la inevitabilidad es una consideración extendida e interesada por diversas razones, entre ellas las económicas. Más allá de los motivos, son importantes los efectos de presumir la inevitabilidad en la aplicación de la tecnología digital, porque desde tal premisa se corre el riesgo de dejar en manos privadas decisiones políticas, sin la participación ciudadana propia de los sistemas democráticos⁴.

De la neutralidad, imparcialidad y objetividad, como elementos propios de los sistemas de inteligencia artificial, se ha llegado a derivar, incluso, su justicia. Con auxilio de las matemáticas, como base de los algoritmos utilizados en los procesos de inteligencia artificial, se pretende trasladar la idea de que los resultados que proporcionan son racionales e imparciales, aportando un conocimiento y tratamiento de la realidad liberados de los condicionantes subjetivos propios del ser humano⁵.

Sin embargo, los algoritmos constituyen una secuencia de instrucciones, ordenada y finita, realizada por personas. Su diseño y desarrollo se realiza desde la perspectiva de sus programadores y de

⁴ Shoshana Zubboff, *La era del capitalismo de la vigilancia. La lucha por un futuro humano frente a las nuevas fronteras del poder*. Trad. por Albino Santos (Barcelona: Paidós, 2020), 458-460.

⁵ Lewis Mumford, *El mito de la máquina. Técnica y evolución humana (1)*. Trad. por Arcadio Rigodón, (Logroño: Pepitas de calabaza, 2013), 329.

quienes se encargan de su realización y puesta en práctica. Personal técnico que impregna a la tecnología de los valores culturales de la sociedad en que funcionan, de los suyos propios y de los intereses particulares de quienes los contratan. De alguna manera, los algoritmos no son sino «opiniones incrustadas en las matemáticas»⁶.

A través de los procesos algorítmicos, usados en la inteligencia artificial, se ordenan, recopilan, seleccionan y etiquetan datos, se eligen variables, se definen problemas, se predicen comportamientos, en definitiva, se atribuye un valor y esto tiene como consecuencia que se ganen o se pierdan posiciones⁷. Por otro lado, hay que ser conscientes que toda elección y atribución de valor lleva aparejado un margen de subjetividad relevante. En este sentido, las nuevas tecnologías se sitúan e impregnan de las condiciones del contexto particular en el que se desarrollan, por tanto, su objetividad y neutralidad es probable que sea una apariencia y aquella atribución de valor esté necesitada de ser desenmascarada y, en su caso, corregida.

Desde la consideración de la infalibilidad tecnológica se extiende la idea de que la inteligencia artificial no se equivoca y la exactitud forma parte de sus atributos. Bajo tal presupuesto, los sistemas de inteligencia artificial adolecen de escasos medios de autocorrección, lo que se evidencia, tanto en la práctica de recogida de datos y entrenamiento, como en el desvío de los objetivos fijados en la programación a medida que la inteligencia artificial adquiere una mayor autonomía de funcionamiento.

Es una realidad que la inteligencia artificial puede acumular una cantidad de datos, e identificar patrones de comportamiento, con una eficacia que supera toda expectativa de conocimiento humano. Sin embargo, la cantidad no es sinónimo de calidad y la abundancia de datos no es equivalente al conocimiento y, mucho menos, es una garantía de exactitud o de imposibles equivocaciones. Por un lado, porque los datos que nutren la inteligencia artificial son aportados por seres humanos y tratados a través de fórmulas matemáticas, introducidas también por personas. Pues bien, los individuos no tenemos entre nuestras cualidades la infalibilidad, ni la imposibilidad de errar. Por lo tanto, si los datos pueden ser incorrectos e incompletos y las fórmulas contener errores, los sistemas de inteligencia artificial no son infalibles⁸.

La dificultad de adoptar una perspectiva crítica en los procesos tecnológicos, atribuible a aquellas falaces presunciones de inevitabilidad, objetividad e infalibilidad, puede agravarse por su carácter inescrutable o, al menos, la dificultad para descifrar su funcionamiento interno. La comprensión del funcionamiento de los sistemas de inteligencia artificial está reservada a expertos, resultando

⁶ Lucía Aragüez Valenzuela, *Hacia la eticidad algorítmica en las relaciones laborales* (Murcia: Laborum, 2024), 71.

⁷ O'Neil, *Armas de destrucción matemática...*, 11,17,23.

⁸ Harari, *Nexus...*, 307-357.

en general incomprensibles para la mayoría de la ciudadanía -incluido sus usuarios y destinatarios-. Ante tal nivel de desconocimiento, el desarrollo de la inteligencia artificial otorga un enorme poder a quienes tienen la posibilidad y capacidad para controlar sus usos, las élites económicas y los tecnócratas⁹. Incluso, en algún momento, su funcionamiento puede quedar al margen de los técnicos, cuando con motivo del aprendizaje profundo los algoritmos actúan a modo de «cajas negras» y sus propios programadores pierden el control sobre su ejecución.

Estas características de la inteligencia artificial, las falaces y las reales, contribuyen, además de minimizar las posiciones críticas respecto a sus usos y aplicaciones, a potenciar su desarrollo en un ámbito de libertad sometido a escasos límites. Lo cual supone un intento de permanecer, en la medida de lo posible, en un terreno desregulado y, en ciertos aspectos, mantenerse al margen de los requerimientos motivados por los derechos humanos afectados, entre ellos, la igualdad de oportunidades.

Respecto a esta problemática no es baladí tener en cuenta que uno de los elementos que más influyen en los procesos algorítmicos y la inteligencia artificial, es la relación, más bien la mala o escasa relación, con la diversidad y el desequilibrio de poder entre sus creadores, quienes lo utilizan en su provecho y los afectados por su uso¹⁰.

Para empezar, los equipos de trabajo en las grandes empresas tecnológicas no cuentan entre sus fortalezas con un personal diverso y abierto a distintas situaciones y sensibilidades. Por el contrario, suelen adolecer de una exigua representatividad de personas de diferentes estratos sociales, género, edad, etnia, raza, diversidad funcional, etc. El trabajo en estas compañías tiende a estar desarrollado por equipos homogéneos, cuyos miembros se ajustan normalmente a un modelo estándar, que se corresponde con quienes están a la vanguardia del desarrollo tecnológico, integrado por varones blancos, jóvenes, plenamente capaces y residentes en territorio urbano¹¹.

Difícilmente puede ser neutral, objetivo e imparcial un procedimiento que, desde sus mismos orígenes, se nutre de criterios homogéneos y carece de diversidad entre sus creadores y, consecuentemente, en su configuración y desarrollo. De ahí que tampoco lo serán sus resultados. Los déficits de representatividad y participación de los diferentes estratos de la población en estos

⁹ Stefano Rodotà, *Vivere la democrazia* (Roma: Laterza, 2018), 138.

¹⁰ Rafael De Asís Roig, *Derecho y tecnologías* (Madrid: Dykinson-Universidad Carlos III de Madrid, 2022), 113.

¹¹ Con respecto al género se ha puesto de manifiesto que el sector tecnológico es, incluso, menos igualitario que el mundo real. Laura Flores Anarte, «Sesgos de género en la inteligencia artificial: el Estado de derecho frente a la discriminación algorítmica por razón de sexo», *Revista Internacional de Pensamiento Político*, vol. 18 (2023): 103.

trabajos tenderán a proyectarse en el diseño de los algoritmos en forma de sesgos y sus posibles efectos discriminatorios. Históricamente los mismos grupos que protagonizan los avances tecnológicos son sus principales beneficiarios, lo que se traduce en desigualdad e injusticia para el resto de las personas que no pertenecen a aquellos estratos sociales. De ahí el racismo, el clasismo, el capacitismo o el machismo, entre otras discriminaciones. Las nuevas tecnologías no lo han inventado, pero lo reproducen y pueden, incluso, amplificarlo.

En conexión con la relación complicada de la inteligencia artificial con la igualdad y la diversidad, está el aprovechamiento económico de los datos. Los datos son actualmente un factor económico de primera magnitud; se habla de los datos como el petróleo de los nuevos tiempos, de la economía del dato y de la economía de la vigilancia. La esencia de la economía de los datos es la diferencia y la parcialidad. Su rentabilidad económica depende, en gran medida, de la parcelación, segmentación y perfilado de los usuarios en red y de explorar sus peculiaridades para distribuir información, bienes o servicios diferenciados.

Para concluir, e incidiendo en esta cuestión, los datos son una fuente económica y una fuente de poder y, más específicamente, de desequilibrio de poder. A efectos de la protección de datos se genera una situación de desequilibrio de poder entre las empresas y organismos que solicitan y almacenan dichos datos y los usuarios que los aportan. Quien proporciona los datos, el usuario de la red, se ve desprovisto de su poder, de su privacidad y autonomía, en favor de quien accede a ellos, los gestiona y administra, pudiendo intervenir en la esfera de actuación de otras personas, para obtener una rentabilidad que va más allá de aspectos puramente económicos¹². En este sentido, todas las personas son vulnerables frente a tal exposición de privacidad. Sin embargo, no es menos cierto que algunas personas, en determinados contextos, tienen más posibilidades de estar expuestos y sufrir mayores daños. Por ejemplo, en relación a la comprensión de la información sobre el tratamiento de los datos o sobre el ejercicio de los derechos de acceso o rectificación, a algunas personas les resultará más difícil la prestación del consentimiento libre y consciente y tendrán más riesgo de sufrir los daños que puedan derivarse de ello¹³.

El *big data*, y el procedimiento algorítmico que sirve para organizar los datos, y configurar realidades conforme a ellos, no puede

¹² Zuboff, *La era del capitalismo de la vigilancia...*, 21, 268, 447. Cristina Monereo Atienza, «Autonomía y vulnerabilidad en la era del capitalismo de la vigilancia. La pervisión de la dimensión humana relacional», *Anuario de Filosofía del Derecho*, vol. XXXVIII (2022): 149-150.

¹³ Julia Anmerman Yebra, «Las personas vulnerables ante el derecho a la protección de datos personales», en *La privacidad en el metaverso, la inteligencia artificial y el big data. Protección de datos y derechos al honor*, coord. por Ángel Acebedo (Madrid: Dykinson, 2022), 53.

considerarse una realidad neutral. Yuval Han señala que «el dataísmo» es «en sí mismo una ideología» que conduce al «totalitarismo digital»¹⁴. Su enorme potencial económico se puede traducir en un reparto de recursos y gestión de las vidas que pone en riesgo los valores democráticos, entre ellos la igualdad¹⁵.

Volviendo a la intencionalidad de ensalzar aquellas características de la inteligencia artificial que abogarían por un uso de estas tecnologías carente de control, se han de tener en cuenta los riesgos más evidentes de su impacto sobre los derechos de la ciudadanía, a fin de proyectar un desarrollo tecnológico que no sea contrario a los valores del Estado de derecho. Estos riesgos son, por un lado, la privacidad, la autonomía y la libertad. De otro lado, la igualdad, la acentuación de discriminaciones preexistentes y la aparición de nuevas formas de discriminación.

4.- DE LAS BRECHAS DIGITALES A LOS SESGOS Y A LA DISCRIMINACIÓN ALGORÍTMICA

Centrados en la incidencia de las nuevas tecnologías y, en particular de la inteligencia artificial, en la igualdad, es preciso tener en cuenta tres elementos fundamentales sobre los que sería necesario actuar para minimizar su impacto negativo y proyectar un desarrollo tecnológico capaz de producir transformaciones en unas sociedades que aspiran a ser más equitativas. Estos factores, cuya incidencia es mayor en los grupos vulnerables, son: las brechas digitales, los sesgos y la discriminación algorítmica.

La vulnerabilidad es una característica universal del ser humano vinculada a su fragilidad¹⁶, que tiene un carácter relacional respecto a elementos sociales y contextuales. Son diversas las circunstancias que condicionan la vivencia de una persona como vulnerable; por ejemplo, la pertenencia a un grupo en situación de desventaja histórica, las necesidades especiales requeridas para desenvolverse en la vida y ejercer los derechos, la posición social, los recursos económicos disponibles, la situación familiar, el lugar de residencia o los apoyos proporcionados por las instituciones públicas. Entre los distintos condicionantes contextuales en el siglo XXI han adquirido una dimensión extraordinaria el uso de la inteligencia artificial, el valor de los datos y la aplicación de los algoritmos a estos datos¹⁷.

¹⁴ Byung-Chul Han, *Psicopolítica. Neoliberalismo y nuevas técnicas de poder*. Trad. por Alfredo Bergés (Barcelona: Herder, 2015), 88.

¹⁵ Cathy O'Neil, *Armas de destrucción matemática...*, 243-269.

¹⁶ Hart se refiere a la vulnerabilidad humana como uno de los elementos que configuran el contenido mínimo del derecho natural. Herbert L. A. Hart, *El concepto de Derecho*. Trad. por Genaro R. Carrió (Buenos Aires: Abeledo-Perrot, 1961), 239-241.

¹⁷ Terrones Rodríguez, «Inteligencia artificial fiable...», 266-267.

Las brechas digitales se refieren a las condiciones desiguales en que algunas personas y grupos acceden, utilizan y obtienen provecho de los recursos tecnológicos. La tecnología se ha convertido en una nueva barrera de la era digital que impide conseguir un desarrollo pleno, inclusivo y sostenible a determinados colectivos, al contribuir a potenciar la pérdida de oportunidades propiciadas por los nuevos recursos que ofrece la técnica. Un nuevo factor de desigualdad que interactúa con otros ya tradicionales, como la edad, el sexo, los recursos económicos, el lugar de residencia, la raza o la discapacidad, potenciando las posibilidades de exclusión, o las dificultades de inclusión, de estos colectivos. A veces las barreras se refieren a ámbitos concretos, como la educación, el empleo, la sanidad, las finanzas o la vivienda, pero siempre están presentes de forma transversal.

Las brechas digitales pueden tener distinto origen. El primero de ellos tiene que ver con las mayores dificultades de acceso material a los recursos tecnológicos, por falta de disponibilidad de dispositivos o por infraestructuras deficientes. La capacidad económica o el lugar de residencia pueden ser un obstáculo para el acceso a la tecnología. La segunda causa de las brechas está relacionada con el uso que se pueda hacer de los recursos disponibles. En este caso, las deficientes motivaciones, habilidades y competencias representan los principales obstáculos para utilizar la técnica en condiciones equitativas. Finalmente, estas barreras conducen a una desigual apropiación y aprovechamiento de las herramientas tecnológicas para la obtención de sus potenciales beneficios en la vida cotidiana, que terminan proyectándose en un menoscabo en el ejercicio y disfrute de los derechos.

Si bien las nuevas tecnologías han pretendido hacer gala de su fácil acceso, de no exigir grandes recursos, ni especiales habilidades y, por ello, de su contribución a la igualdad, lo cierto es que estas consideraciones están lejos de ajustarse a la realidad. El universo digital está abierto a todas las personas, pero pese a ello se ha creado una «clase baja digital», compuesta por personas con pocos recursos económicos, los pobres, quienes carecen de una adecuada alfabetización digital, que disponen -en el mejor de los casos- de equipos obsoletos, que acceden a la red con conexiones lentas y poco fiables¹⁸. Ahora bien, no solo las personas con escasos recursos económicos se encuentran con ese tipo de obstáculos.

Como punto de partida, en relación a la tecnología digital, actualmente se distinguen dos grupos de personas. Los nativos digitales, que han nacido en el periodo de creación y expansión de las nuevas tecnologías, y los inmigrantes digitales quienes, con más o menos fortuna, se han ido incorporando a este medio. Las personas de

¹⁸ Informe del Relator Especial *sobre la extrema pobreza y los derechos humanos* A/74/493, de 11 de octubre de 2019: 16, último acceso 22 de marzo 2025, https://digitallibrary.un.org/record/3834146/files/A_74_493-ES.pdf

edad avanzada encuentran serios problemas para adaptarse al medio, en algunos aspectos también las personas con discapacidad¹⁹. El incremento de la longevidad, y su relación con la discapacidad, hacen de la edad un factor importante para explorar las posibilidades de la tecnología digital en la superación de barreras, en la promoción de la autonomía, en la participación y la inclusión de estos colectivos en la vida social, en el bienestar y la calidad de vida de los grupos afectados y de la sociedad en general²⁰. Diversos estudios ponen de manifiesto que estas barreras, traducidas en brechas digitales, abocan en sustanciales variaciones en cuanto a los niveles de lectoescritura y de alfabetización digital, que repercuten en la posibilidad, o no, de realizar actividades *online* de manera directa, autónoma y segura²¹.

El bajo nivel de alfabetización y competencia digital, especialmente en las personas mayores, se acompaña de la falta aceptación y resistencia a los bruscos cambios que en su vida pretenden introducir unas herramientas tecnológicas desarrolladas, por otro lado, sin su implicación, ni participación en su diseño²². Al respecto, tan importante como afrontar las barreras económicas, y las relativas a la capacidad o facilidad para manejarse con los aparatos y mecanismos digitales, son las condiciones de diseño y programación de los dispositivos, medios y recursos tecnológicos, a fin de facilitar su adaptación y aceptación por todos los posibles usuarios y usuarias.

¹⁹ Los colectivos en que la brecha digital es más aguda son los adultos mayores y las personas con discapacidad. Entre otras barreras que dificultan el acceso a la tecnología pueden encontrarse asistentes de voz inteligente que no reconocen la voz de personas con dificultades de habla, o *captchas* que no son capaces de ser superados por determinadas personas o sistemas de reconocimiento facial que incrementan los errores en rostros con determinadas peculiaridades derivadas de alguna discapacidad. Héctor Álvarez García, «La discriminación algorítmica por razón de discapacidad», *Revista Internacional de Pensamiento Político*, vol. 18 (2023): 23, 30.

²⁰ Carmen Llorente-Barroso, Eva Muñoz-Paniagua y Francisco García García, «Innovación digital en el ecosistema de la *silver economy* para responder a las discapacidades derivadas del envejecimiento», en *Vulnerabilidad digital: desafíos y amenazas de la sociedad hiperconectada*, ed. por Rebeca Suárez Álvarez, Miguel Ángel Martín Córdoba y Luis Manuel Fernández Martínez (Madrid: Dykinson, 2023), 188-197.

²¹ Gloria Esperanza Álvarez Ramírez, *La protección de las personas consumidoras vulnerables con discapacidad frente a la brecha digital* (Madrid: Cinca, 2024), 62.

²² Raquel Valle Escolano, «Inteligencia artificial y derechos de las personas con discapacidad: el poder de los algoritmos», *Revista Española de Discapacidad* 11(1) (2023): 17-18, doi: <https://doi.org/10.5569/2340-5140.11.01.01>. La escasa participación de las personas con discapacidad en la elaboración y la toma de decisiones de los sistemas de inteligencia artificial; el diseño, que se nutre de datos estereotipados; los sistemas basados en modelos de normalización, sin tener en cuenta la diversidad; y el propio uso de la inteligencia artificial para identificar y discriminar a las personas con discapacidad, son los riesgos de discriminación que identifica Francisco Javier Bariffi, «Inteligencia artificial, derechos humanos y discapacidad ¿reflejo de los prejuicios humanos u oportunidad del transhumanismo» en *Nuevas fronteras del derecho y de la discapacidad*, 2, ed. por Rafael de Lorenzo García y Luis Cayo Pérez Bueno (Cizur Menor: Aranzadi, 2021), 121.

Estas fases del desarrollo tecnológico están concebidas pensando en un modelo estándar y normalizado de personas, como consecuencia, quienes no se ajustan al mismo, encuentran dificultades para poder hacer uso de los recursos tecnológicos disponibles. Por otro lado, el diseño de la tecnología considerando a un tipo concreto de sujetos, terminan confirmando y proyectando en el imaginario colectivo lo que debe suponer un cuerpo y mente normal y, con ello, los prejuicios hacia los grupos de personas que no se adecúan al mismo.

Junto con la edad y la capacidad, otras brechas digitales tienen como causa el género, en cuyo caso resultan desfavorecidas las mujeres; el lugar de residencia, que privilegia a quienes habitan un ámbito urbano, frente quienes lo hacen en territorio rural; o la raza, que coloca en situación de vulnerabilidad a las personas de color.

Las barreras que dificultan el acceso, la comprensión y el aprovechamiento de los recursos digitales generan desconfianza y falta de seguridad en el entorno digital. Las competencias digitales son necesarias para el aprendizaje permanente, la realización personal, mantener un estilo de vida saludable y sostenible, participar en el entorno profesional y construir una ciudadanía activa e inclusiva.

El mundo digital es hoy parte del mundo real, es el entorno en que se proyectan múltiples aspectos de la vida personal, social, cultural, educativa, laboral, sanitaria y de comunicación. Como consecuencia, para poder participar activamente en todos los ámbitos de la vida es fundamental la accesibilidad en condiciones de igualdad al medio digital de todas las personas. Negar u obstaculizar el acceso a la red es levantar muros virtuales y establecer barreras electrónicas que implican, en la realidad de este ciudadano digital, lesionar sus derechos fundamentales y mermar sus posibilidades de participar activamente en la vida cultural, social y política²³. En el estado del desarrollo de la tecnología, y sus diversos usos, la accesibilidad es un derecho necesario para acompañar la metamorfosis de una ciudadanía analógica a una ciudadanía digital. En estos términos, el acceso a internet, en condiciones de igualdad, se convierte en un bien público de primera magnitud para el ejercicio de los derechos de la ciudadanía y el buen funcionamiento del sistema democrático²⁴. La accesibilidad, como condición de posibilidad y como derecho autónomo, puede extenderse a todo el sistema de derechos y a todas las personas. Como derecho singular y autónomo se proyecta como un derecho prestacional, lo que supone el deber de los poderes públicos de establecer las medidas oportunas para que toda la ciudadanía conozca y tenga posibilidad de acceder a las nuevas tecnologías de la comunicación. Como condición de posibilidad del ejercicio de otros

²³ Tommaso Edoardo Frosini, *Libertad, igualdad, internet* (Ciudad de México: Tirant lo blanch, 2019), 53.

²⁴ Isabel Victoria Lucena Cid, «El derecho de acceso a Internet y el fortalecimiento de la democracia», *Revista Internacional de Pensamiento Político*, vol. 9 (2014): 383-398.

derechos, formaría parte de su contenido esencial, en tanto, los derechos carecerían de una real y amplia eficacia sin su garantía²⁵.

La accesibilidad, en su sentido más amplio, es un derecho que se concreta y desarrolla en el artículo 9 de la *Convención Internacional de los Derechos de las Personas con Discapacidad*, incluyendo el acceso a los nuevos sistemas y tecnologías de información y comunicación. Especialmente en relación a este colectivo los dispositivos y avances tecnológicos tienen potencialmente un gran poder igualador y favorecedor de la inclusión social, entre otros ejemplos son de gran utilidad los robots de asistencia personal, los equipos de movilidad, los dispositivos para la audición, los traductores virtuales o los medios para teletrabajar.

El Informe del Relator Especial de las Naciones Unidas sobre los derechos de las personas con discapacidad al Consejo de Derechos Humanos A/HRC/49/52, de 28 de diciembre de 2021, trata específicamente sobre la inteligencia artificial y los derechos de las personas con discapacidad. Aborda las repercusiones de la inteligencia artificial en las personas con discapacidad; los derechos y obligaciones que deben cumplirse en su desarrollo e implementación, de acuerdo con la *Convención Internacional de los derechos de las personas con discapacidad*; sus aplicaciones y sus riesgos para el colectivo. Finalmente, termina haciendo una serie de recomendaciones, con el objetivo de conseguir que los procesos de inteligencia artificial sean inclusivos, diversos y equitativos²⁶.

En atención a esta nueva realidad, *La Recomendación sobre la Ética de la Inteligencia Artificial* de la UNESCO, de 23 de noviembre de 2021, señala, entre las preocupaciones éticas fundamentales del desarrollo de la inteligencia artificial, la aparición de divisiones sociales y económicas generadas por las nuevas brechas y las desigualdades que tienen su origen en los procesos digitales. Se entiende por ello que los sistemas de inteligencia artificial suponen una amenaza para la diversidad y un factor de exclusión. En su apartado 28 aboga por un enfoque inclusivo de los beneficios de la inteligencia artificial, para lo cual habrá de ser accesible, asequible y tener en cuenta las necesidades específicas de distintos colectivos por razones de edad, cultura, lengua, capacidades diversas, género, nacionalidad, personas desfavorecidas, marginadas y vulnerables o en situación de vulnerabilidad²⁷.

²⁵ Ambas dimensiones se corresponden con el sentido débil y fuerte de la accesibilidad, que ha de estar presente en todo enfoque de derechos y, también, en el desarrollo de la inteligencia artificial. Rafael De Asís Roig, «De nuevo sobre la inteligencia artificial y los derechos humanos», *Derechos y Libertades*, n.º 51 (2024): 37-40.

²⁶ El informe está disponible en: <https://docs.un.org/es/A/HRC/49/52>, último acceso el 25 de marzo de 2025.

²⁷ <https://unesdoc.unesco.org/ark:/48223/pf0000381133/PDF/381133eng.pdf.multi.page=62>, último acceso el 25 de marzo de 2025.

La tecnología, por tanto, no necesariamente favorece la inclusión de todas las personas, más bien parece que continúa con los favoritismos tradicionales por el género masculino, la raza blanca, la clase media y alta, las personas sin discapacidad y las que habitan en núcleos urbanos. Con todo, la técnica ha traspasado, desde el ámbito analógico hasta el digital, los problemas de inclusión e igualdad, entre otros motivos, porque las brechas digitales se han convertido en una nueva fuente de desigualdad en el siglo XXI²⁸.

Las brechas digitales son una de las causas de la existencia de sesgos en los sistemas de inteligencia artificial. Los datos, que hacen posible la dinámica algorítmica, son generados por la actividad en red de sus usuarios y usuarias. Si determinados colectivos encuentran barreras de acceso, y tienen poca interacción tecnológica, se produce una desproporción de la información referida a sus integrantes, en relación con otros grupos que no padecen aquellas dificultades²⁹. En este sentido, existen «zonas oscuras o sombras» en los datos, los de las personas, grupos o comunidades ignorados para su obtención. Personas que viven al margen de la red, o actúan en ella con poca intensidad. Se han señalado ya algunas circunstancias, como las relativas al género, la edad, la discapacidad, la pobreza o los entornos rurales; otras tienen que ver con el estilo de vida o, simplemente, porque algunas personas resultan menos interesantes al mercado o negocio de los datos³⁰.

La deficiente participación de estos colectivos en todo el proceso de la inteligencia artificial, y sus aplicaciones concretas, provoca que los criterios que se tienen por generales, en realidad son parciales al referirse solo a las personas de quienes se han extraído los datos.

La escasa participación de determinados grupos en la red no es la única causa de las deficiencias en los datos. Cualquier recogida de datos y su tratamiento matemático opera a través de una simplificación sobre la complejidad de la realidad, con la toma de decisiones previas sobre aquello que es importante, relevante y debe ser considerado como una prioridad. Aunque la inteligencia artificial puede operar sobre

²⁸ Judy Wajcman, *El tecnofeminismo*. Trad. por Magalí Martínez Solimán (Valencia: Cátedra, 2004), 14, 97-98.

²⁹ En el caso de las mujeres, la brecha digital de género hace que haya menos datos sobre ellas y, por lo tanto, que estén infrarrepresentadas en el *big data*. Laura Flores Anarte, «Sesgos de género en la inteligencia artificial: el Estado de derecho frente a la discriminación algorítmica por razón de sexo», 102.

³⁰ Algunos colectivos son menos atractivos desde el punto de vista de muchos negocios desarrollados en la red por tener menos capacidad económica. Por tal motivo, se tienden a disminuir considerablemente las ofertas a estas personas dirigidas y, con ello, la mayor dificultad de acceder a bienes y servicios. Lo que hace que la inteligencia artificial contribuya a ampliar las diferencias económicas y sociales en grupos ya desfavorecidos. Jorge Castellanos Claramunt, «Human rights, vulnerability and artificial intelligence: an analysis in constitutional perspective», *Deusto Journal of Human Rights* 14 (2024): 41, 44, doi: <https://doi.org/10.18543/djhr.3187>

una cantidad de datos inaccesibles al ser humano, es imposible que pueda aborarlos todos. A mayor abundamiento, en la obtención, elección y tratamiento de datos hay intervención humana y se incide sobre seres humanos. Si en las personas existen sesgos, también los habrá en cualquier proceso en que participen.

Los sesgos pueden producirse al traducir a lenguaje informático el problema que se quiere resolver en el proceso computacional, en la recopilación de los datos de entrenamiento, en su etiquetado, en la selección de las características que han de tenerse en cuenta, en el peso de los distintos factores, en las variables de aproximación o en la toma de decisiones sobre la base del modelo y sus datos³¹. Pueden derivarse de una selección de datos que sean, o no, representativos de la realidad. Es decir, se puede tratar de datos incompletos, por no reflejar la diversidad de personas y circunstancias que existen en la sociedad. Pero también puede que sean reflejo de una desventaja histórico-social que, efectivamente, se sigue produciendo en la realidad y, por tanto, la representa. Los sesgos son entonces preexistentes, propios de las instituciones sociales, prácticas y actitudes de una sociedad y cultura determinadas.

Esos datos sesgados pueden ser introducidos por el programador de forma inconsciente, al no tener en cuenta variables más precisas, tratar con datos incompletos o erróneos. Pero también puede hacerlo de forma consciente, a modo de respaldo o reflejo de los prejuicios existentes e, incluso, puede que no procedan directamente del programador, sino que se deriven de la propia retroalimentación del algoritmo

Existen también sesgos originados por problemas técnicos, como un uso inadecuado de las herramientas tecnológicas, un fallo en el aprendizaje del algoritmo o una errónea formulación matemática. Puede que el sesgo sea emergente, es decir, que se produzca cuando el sistema interactúa con la realidad a la que se aplica la programación, motivado por los cambios sociales y los valores culturales de la población. Con respecto a las decisiones o resultados, los sesgos se producen porque no se han comprobado las posibles consecuencias derivadas de la aplicación a la realidad del algoritmo, o porque, pese a que se percibieron, no fueron corregidas³².

³¹ Solon Barocas y Andrew D. Selbst, «Big Data's Disparate Impact», *California Law Review* 104 (2016): 675, doi: <http://dx.doi.org/10.15799/Z38BG31>.

³² Batya Friedman y Helen Nissebaum, «Bias in Computer Systems», *ACH Transactions on Information Systems*, 14/3 (1996): 333-336, doi: <https://doi.org/10.1145/230538.230561>. Sesgos de datos, algorítmicos, emergentes, de confirmación, de sobreconfianza, de falta de diversidad, de retroalimentación de contexto o de explicación son recogidos por: Marco Emilio Sánchez Acevedo, «Buena administración algorítmica y debido proceso frente a los sesgos», en *Algoritmos abiertos y que no discrimine en el sector público*, ed. por Lorenzo Cotino Hueso y Jorge Castellanos Claramunt (Valencia: Tirant lo blanch, 2023), 88-89.

Si, como se ha señalado, los datos necesariamente operan una reducción sobre fenómenos, objetos o comportamientos del mundo real, con respecto a los grupos desfavorecidos la merma puede llegar hasta la ignorancia total. Configurar los criterios extraídos a partir de personas con unas características determinadas como patrones de comportamiento generales para todas las personas, impide que estos parámetros se ajusten a aquellos colectivos que han sido ignorados y, por tanto, a sus tendencias, aspiraciones, preferencias, deseos o comportamientos. Las decisiones entonces se inclinarán a favor de algunos grupos concretos y en perjuicio de otros. En definitiva, de datos sesgados se derivarán efectos prejuiciosos y perjudiciales para determinadas personas y grupos; decisiones erradas, en el mejor de los casos, discriminatorias en algunos otros³³.

5.- PECULIARIDADES Y DIFICULTADES PARA PREVER, DETECTAR Y CORREGIR LA DISCRIMINACIÓN ALGORÍTMICA

Los problemas de determinados colectivos para acceder a bienes y servicios, los sesgos en la toma de decisiones, y sus consecuencias en términos de discriminación, no se producen exclusivamente por la intervención de la inteligencia artificial, el procesamiento de datos y las fórmulas algorítmicas a ellos aplicadas. Ahora bien, la utilización de estos sistemas presenta algunas peculiaridades que se traducen en dificultades para prever, detectar y corregir la discriminación.

En primer lugar, las predicciones y toma de decisiones basadas en datos implican la proyección del pasado al presente y al futuro. Esto significa que lo venidero puede quedar lastrado por hechos y acontecimientos anteriores, por una sobreabundancia de datos históricos que dimensionan, en exceso, factores que han perjudicado tradicionalmente a determinados grupos. Estos colectivos ignorados en el pasado estarán necesariamente subrepresentados en los datos recopilados y, en consecuencia, en toda decisión basada en ellos.

La inteligencia artificial, en este sentido, camina de espaldas al futuro, entraña un impedimento para que irrumpa lo nuevo, para avanzar hacia «lo intransitado», hacia aquello que puede impulsar un cambio sobre lo ya acaecido. El peso del pasado puede suponer un serio obstáculo para establecer medidas que compensen, favorezcan y generen expectativas a aquellos sujetos tradicionalmente privados de oportunidades, quienes han padecido discriminaciones.

A través de procesos de inteligencia artificial se puede calcular y decidir en base a probabilidades, elegir entre distintas opciones predeterminadas, pero no generar otras nuevas propias del quehacer de la inteligencia humana, capaz de construir pensamiento original³⁴.

³³ Barocas y Selbst, «Big Data's Disparate Impact», 684-688.

³⁴ Byung-Chul Han *No-cosas. Quiebras del mundo de hoy*. Trad. por Joaquín Chamarro Mielke (Madrid: Taurus, 2022), 59.

Porque en estos procesos, no hay, en principio, lugar para juicios de valor individuales que puedan rectificar y modificar la experiencia pasada y permitan proyectarse hacia un presente y un futuro más igualitario.

Todo lo contrario, la repetición de patrones del pasado, replicados de forma automatizada, puede reforzar, reproducir, perpetuar y codificar los sesgos ya existentes, seguir condenando a la invisibilidad a determinados grupos, reafirmar su estigmatización, revalidar los estereotipos existentes y, finalmente, acentuar la discriminación que sobre ellos pesaba. Al respecto se ha señalado que mientras los sesgos discriminatorios en las decisiones humanas incrementan los perjuicios «a modo sumatorio», las decisiones automatizadas lo hacen «en términos de multiplicación»³⁵.

Por lo tanto, los sistemas de inteligencia artificial con sus fórmulas matemáticas de base, más que corregir, tienden a agravar las divisiones y desigualdades existentes, complicando extraordinariamente la posibilidad de romper ciertos estereotipos o barreras, que requieren dejar el pasado y mirar al futuro con perspectiva y, especialmente, precisan de una visión transformadora más propia, para estos casos, de la inteligencia humana. Frente a «la exacerbación del solucionismo tecnológico» hay que establecer estrategias humanas³⁶. La auténtica revolución digital no sería conseguir la eficiencia en términos de mercado sino en términos humanos, conseguir los mejores servicios y cuidados para quienes más lo necesitan. Los sistemas de inteligencia artificial han de ser un apoyo, pero no deben relegar lo creativo e innovador del ser humano.

Otra de las características de los procesos de carácter técnico que puede contribuir a codificar, cronificar o consolidar los sesgos es la presunción de objetividad. Es el llamado «sesgo pro-máquina» o sesgo de automatización, por el que se tiende a mantener la creencia de que los resultados de los procesos matemáticos son objetivos, carentes de elementos subjetivos y valorativos y, por ello, tienden a ser aceptados de manera acrítica³⁷. Desde este planteamiento se

³⁵ José Fernando Lousada Arochena, «Inteligencia artificial y sesgos discriminatorios: ¿Es necesario un nuevo concepto de discriminación algorítmica?», *IgualdadES* 11(2024):106, doi: <http://doi.org/10.18042/cepc/IgdES.11.04>

³⁶ Terrones Rodríguez, «Inteligencia artificial fiable...», 270. Al menos abordar con una mayor amplitud el impacto de la inteligencia artificial, lo que requeriría abandonar el «tecnosolucionismo» para adoptar un enfoque más social. Cristina San Miguel Caso, «Reflexiones sobre la prevención y mitigación de los sesgos en los sistemas inteligentes», en *Más allá de la justicia: nuevos horizontes del Derecho procesal*, dir. por Ana Sánchez Rubio, coord. por Paloma Arrabal Platero y José Caro Catalán (Valencia: Tirant lo blanch, 2024), 236.

³⁷ Alba Soriano Aranz, «La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicaciones de inteligencia artificial», en *Nuevas normatividades: Inteligencia artificial, Derecho y género*, ed. por Pablo Raúl Bonorino Ramírez, Rafael Fernández Acevedo y Patricia Valcárcel Fernández (Cizur Menor-Navarra: Aranzadi, 2021), 69.

acaban enmascarando como neutrales e imparciales cuestiones que pueden llegar a tener un marcado carácter ideológico e, incluso, discriminatorio y sectario.

Entre las peculiaridades de los sistemas de inteligencia artificial, que dificultan abordar sus posibles efectos discriminatorios, está también la falta de transparencia y la complejidad -en ocasiones, imposibilidad- para explicar sus procedimientos de forma fácilmente comprensible para los afectados. La recogida, análisis y tratamiento de datos se utilizan para facilitar gestiones dirigidas a conceder determinados recursos y beneficios, o bien para denegarlos. En virtud del cumplimiento de finalidades tan importantes para la ciudadanía, lo lógico, justo y democrático, sería que se pudiera conocer, y comprender, cómo se llega a aquellas decisiones. Sin embargo, el funcionamiento de los algoritmos a este respecto es difícilmente comprensible para el ojo humano común. Por tal motivo, O'Neil, señala que la opacidad, la escala y el daño conforman las «armas de destrucción matemática» de los modelos de decisión impulsados por los algoritmos³⁸.

La opacidad supone un obstáculo determinante para identificar y solucionar los problemas de discriminación algorítmica. Por un lado, la estructura del algoritmo y los datos de entrenamiento pueden estar protegidos por la propiedad intelectual y el secreto industrial, como consecuencia ser de acceso restringido. Por otro lado, los algoritmos de aprendizaje automático operan como cajas negras que no permiten una trazabilidad y explicación completa de sus resultados. Habida cuenta de estas restricciones sobre el conocimiento del algoritmo, es posible que se ignoren los datos que se utilizan, el sistema de correlaciones establecido, los códigos o el valor de los distintos factores que han podido ser determinantes para alcanzar un determinado objetivo. Si es así, ¿cómo conocer si había una oferta de trabajo posible para nuestro perfil y no nos ha llegado? ¿Cuál es el motivo? ¿Cómo defendernos de un algoritmo que nos atribuye un alto riesgo de no devolver un préstamo? ¿Por qué razón la prima de nuestro seguro médico es sustancialmente más caro que el de otra persona de condiciones de salud similares?... Es claro que, si no se cuenta con la información necesaria, la posibilidad de acceder y tener las explicaciones adecuadas para comprender el funcionamiento del algoritmo, sus objetivos y sus resultados, será difícil detectar la existencia de una posible discriminación.

A la dificultad de conocer si se está produciendo o no una conducta discriminatoria, en caso de que existan indicios de que así sea, hay que añadir la complejidad de probarlo. La prueba de la discriminación es complicada en todos los supuestos, si confluyen o interactúan diversos elementos de discriminación, lo será aún más. En el caso de la inteligencia artificial, la cantidad de datos que maneja

³⁸ O'Neil, C *Armas de destrucción matemática...*, 43.

implica, con un alto grado de probabilidad, la convergencia simultánea de múltiples factores de discriminación, así como, probablemente, el desconocimiento de cuántos y cuáles y del peso específico de cada uno de ellos. Todas estas circunstancias entorpecen significativamente la posibilidad de asociar el prejuicio y perjuicio a algún elemento concreto.

Los problemas técnicos, tanto los relativos a las limitaciones, como a la complejidad, son causantes de algunas de las particularidades de la discriminación algorítmica. Desde esta perspectiva, pueden considerarse la dificultad de los procesos de inteligencia para valorar globalmente todos los factores que influyen en sus predicciones y decisiones, los errores de correlación o los fallos en las formulaciones matemáticas³⁹.

Para afrontar el problema de la discriminación algorítmica, junto al lastre del pasado, la falta de transparencia y los problemas técnicos, es preciso ser consciente de que hay un negocio de la inteligencia artificial con sus propios objetivos e intenciones. Presidida por la necesidad de rentabilidad y eficiencia, sobre esta industria pesa la sospecha de su despreocupación por la justicia y, por ende, la carencia de esfuerzos por garantizar la accesibilidad, la diversidad y la inclusión.

Los ejemplos de brechas digitales, los sesgos, sus efectos discriminatorios, y cómo han afectado a los derechos, son numerosos y bien conocidos. Los filtros informativos y la publicidad conductual a través del *microtargeting* están extendidos en las ofertas de publicidad en línea y en los motores de búsqueda. No es el usuario de la red quien elige y decide sobre los contenidos que desea recibir, sino que es el algoritmo quien determina, según su historial de búsquedas y visitas a páginas web, operaciones comerciales realizadas e intereses expresados en red, quien predice y oferta, condicionando así lo que deberían ser elecciones personales libres. A través de estas técnicas los servidores de internet priorizan y bloquean informaciones de forma personalizada, de tal manera que los mismos bienes y servicios no aparecen disponibles para todas las personas. La selección de a quien llegan las ofertas tiene como efecto la exclusión a determinados sujetos de oportunidades laborales, educativas, comerciales, crediticias o culturales⁴⁰.

En el ámbito laboral tuvo una enorme repercusión el algoritmo de contratación que *Amazon* diseñó en 2014 y retiró en 2018 por

³⁹ Junto con la invisibilidad e intensidad, la complejidad técnica es puesta de manifiesto por María José Añón Roig, «Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos», *Derechos y Libertades*, n.º 47 (2022): 36. Por su parte, las limitaciones técnicas son señaladas por Jose Carlos Fernández Rozas, «Ética, desafíos y riesgos de acceso a la justicia algorítmica», *Deusto Journal of Human Rights* 14 (2024): 209, doi: <https://doi.org/10.18543/djhr.3195>

⁴⁰ Eli Pariser, *El filtro de la burbuja. Cómo la red decide lo que leemos y lo que pensamos*. Trad. por Mercedes Vaquero (Barcelona: Taurus, 2017), 18, 116, 131-133.

discriminatorio hacia las mujeres. El algoritmo fue entrenado para detectar patrones en las contrataciones de los últimos 10 años y aprendió que en los puestos técnicos los varones eran preferentes a las mujeres, porque la mayoría de los contratados en ese tiempo, y para este tipo de puesto de trabajo, eran varones. De estos datos el algoritmo infirió que el perfil de mujer no era adecuado y no propuso la contratación de mujeres, pese a que el sexo no era una variable introducida por los programadores.

El caso del algoritmo Frank, utilizado por *Deliveroo*, clasificó la fiabilidad de sus repartidores por su disponibilidad en el trabajo, sin distinguir las causas de sus ausencias o retrasos. De esta manera, penalizó a las personas que estaban de baja laboral, que ejercieron su derecho de huelga o su derecho a conciliar vida laboral y familiar⁴¹.

Sobre las garantías procesales, ha sido analizado, comentado con profusión y sometido a un proceso judicial en el Tribunal Supremo de Wisconsin, *State v. Loomis*, (2016) el programa COMPAS del sistema judicial de EE. UU (*Correctional Offender Management Profiling for Alternative Sanctions*). El sistema evalúa el riesgo de reincidencia entre los acusados, sirviendo de auxilio a los jueces para dictar sentencia, establecer fianzas y conceder libertades condicionales. En 2016 una investigación puso de manifiesto que la imparcialidad de las decisiones judiciales, apoyados por COMPAS, estaba comprometida, entre otros, por el sesgo racial. El estudio señalaba que las personas negras calificadas por el programa de alto riesgo tenían casi el doble de posibilidades de no reincidir que los blancos con la misma calificación; por el contrario, las personas blancas etiquetadas de bajo riesgo tenían más probabilidades de reincidencia que las negras igualmente clasificadas. Se ponía así en evidencia que el sistema judicial norteamericano, históricamente aquejado de discriminación racial, tendía a perpetuar la misma discriminación con los sistemas de inteligencia artificial. El riesgo existía, no obstante, el Tribunal Supremo aceptó la constitucionalidad del programa, si bien consideró preciso adoptar algunas cautelas, como la prohibición de que la decisión judicial estuviera basada exclusivamente en su predicción y la exigencia de que tuviera un carácter meramente orientativo, de auxilio judicial, no imperativo⁴².

En el terreno policial, en el uso de sistemas de vigilancia policial predictiva se han detectado también sesgos raciales y cierta tendencia a la criminalización de determinados colectivos. Uno de los programas informáticos más conocidos es *PredPol* (*Predictive Policing*) creado por

⁴¹ El Tribunal Ordinario de Bolonia en Sentencia de 31 de diciembre de 2020, condena a la empresa por discriminación indirecta contra los trabajadores, al no tratar de forma diferenciada a circunstancias distintas y justificadas en el ejercicio de derechos reconocidos y protegidos por el ordenamiento jurídico italiano.

⁴² Lucía Martínez Garay, «Peligrosidad, algoritmo y due process: el caso *State v. Loomis*», *UNED, Revista de Derecho penal y criminología* n.º 20 (2018): 485-502, doi: <https://doi.org/10.5944/rdpc.20.2018.26484>

la empresa *Geolítica*. Se trata de un algoritmo de aprendizaje automático para predecir dónde y cuándo es más probable que ocurran delitos. Los datos que utiliza se refieren al tipo de delito, la ubicación, la fecha y la hora. Su objetivo es hacer un mapa de actuación policial en determinadas zonas para aumentar la vigilancia, organizando patrullas que puedan disuadir de la comisión del delito y reducir las tasas de criminalidad. Su eficacia contrastada, no impide que se haya puesto de manifiesto que, al tener en cuenta principalmente determinados delitos -como hurtos, robos, alteraciones del orden u otros que impliquen uso de violencia física- y no otros tipos delictivos -como estafas, delitos fiscales o financieros-, la prevención y vigilancia se centra en determinadas zonas de menor renta, con población perteneciente a minorías raciales y étnicas. Al aumentar la presión policial, se producen más detenciones a incorporar a las bases de datos que alimenta el algoritmo, con ello se incrementa la vigilancia y la estigmatización de algunos colectivos, cuyos hogares se concentran en determinadas zonas de las ciudades⁴³.

En ayudas sociales, el programa *Bosco* en España se utilizó para otorgar el bono social eléctrico, destinado a personas en situación de vulnerabilidad, en concreto, a familias numerosas, a quienes dispongan de rentas bajas y a beneficiarios de pensiones mínimas de incapacidad o de jubilación, que no tuvieran otros ingresos. El programa rechazó las solicitudes de las viudas, al no poder ser incluidas entre los y las pensionistas, ni tampoco por el nivel de renta. La Fundación *Civio* solicitó al Consejo de transparencia y buen gobierno información sobre el programa y el código fuente e interpuso un recurso ante la Audiencia Nacional y el Tribunal Supremo quien decidirá, en última instancia, sobre si la propiedad intelectual justifica la opacidad del programa y los límites al acceso a la información, específicamente, cuando se trata de una ayuda social de carácter público.

Otro caso de gran impacto es el de SyRI (*Systeem Risico Indicatie*) utilizado en los Países Bajos y declarado ilegal por el Tribunal de La Haya. El programa evaluaba el riesgo de fraude a la seguridad social y a la hacienda del gobierno neerlandés, en las personas que solicitaron beneficios sociales, llegando a bloquear la posibilidad de acceder a estas ayudas a personas que tenían derecho a ellas en situación de vulnerabilidad socioeconómica. La violación del derecho a la vida privada, la falta de transparencia e información a los afectados, la discriminación y la estigmatización injustificada, fueron los motivos

⁴³ *PredPol* es uno de los programas líderes a nivel mundial, utilizado tanto en EE. UU como en Reino Unido. También Alemania, Francia, Italia y Países Bajos cuentan con cierta experiencia en estos métodos de análisis criminal. Virginia Cinelli y Alberto Manrique Gan, «El uso de programas de análisis predictivo en la inteligencia policial: una comparativa europea», *Revista de Estudios en Seguridad Internacional*, vol. 5, n.º 2 (2019): 8, doi: <http://dx.doi.org/10.18847/1.10.1>. Sobre su aplicación en España, Jose Luís González-Álvarez, Jorge Santos-Hermoso y Miguel Camacho-Collado, «Policía predictiva en España: aplicaciones y retos futuros», *Behavior & Law Journal*, vol. 6, n.º 1 (2020): 26-41, doi: <https://doi.org/10.47442/blj.v6.i1.75>

de la declaración de ilegalidad por parte del Tribunal. El programa utilizaba datos de todas las administraciones públicas, sin informar a los usuarios, a fin de detectar elementos comunes en los defraudadores y ponerlos en relación con las personas solicitantes de ayudas⁴⁴.

También hay ejemplos en el ámbito de la salud. En el año 2019 la Revista *Science* destapó el sesgo racial de un algoritmo utilizado en varios hospitales del sistema de salud de EE. UU. El programa identificaba a personas con ciertas afecciones de salud y riesgo de agravarse, a fin de diseñar un sistema preventivo que, a la vez, pudiera reducir en el futuro el gasto sanitario de estos enfermos. En la investigación se descubrió que las personas negras etiquetadas como de alto riesgo empeoraba la salud a un ritmo mayor que las blancas. El motivo de este agravamiento era que uno de los indicadores que se utilizó, para establecer las medidas preventivas, era el coste en salud de los pacientes. Del menor gasto en salud de las personas de color se dedujo que sus condiciones de salud eran mejores y, por ello, el sistema de prevención invirtió menos en ellos⁴⁵.

Diversos sistemas de reconocimiento facial, basados en datos biométricos, han tenido dificultades para identificar correctamente a determinados grupos de personas en función de raza, etnia o sexo, por utilizar bases de datos poco representativas de la diversidad humana. Circunstancia que puede arrojar discriminaciones varias, en áreas como la seguridad, la vigilancia o la identificación. Otro caso, citado frecuentemente como reproductor de sesgos, hace referencia a los asistentes de voz, o mejor a las asistentes de voz, por transmitir una imagen servicial de la mujer, que contribuye a reforzar y perpetuar los estereotipos de género⁴⁶.

Tampoco los ámbitos financiero y crediticio son indiferentes a los sesgos, al haber excluido o perjudicado, en no pocas ocasiones, el acceso al crédito a determinados colectivos. En 2019 Hansson, un usuario de la tarjeta *Apple Card*, denunció en su cuenta de *Twitter* la oferta de una línea de crédito veinte veces mayor que la que se le

⁴⁴ Lorenzo Cotino Hueso, «SyRY, ¿a quién sancionó? Garantías frente al uso de la inteligencia artificial y decisiones automatizadas en el sector público y la Sentencia holandesa de febrero de 2020», *La Ley Privacidad*, n.º 4 (2020): 1-10.

⁴⁵ Ziad Obermeyer, Brian Powers, Christine Vogeli y Sendhil Mullainathan, «Dissecting racial bias in an algorithm used to manage the health of populations», *Science*, vol. 366, n.º 6464 (2019): 447-453, doi: [10.1126/science.aaxx2342](https://doi.org/10.1126/science.aaxx2342)

⁴⁶ Sobre la potencialidad de los sistemas de inteligencia artificial para reforzar estereotipos es un ejemplo muy representativo el de las asistentes de voz. La asistente de Google, Siri, Alexa o Cortana proyectan una imagen servicial de la mujer e, incluso, tolerante frente a actitudes verbales abusivas, como bien figura en el informe de la UNESCO de 2019 titulado *I'd blush if i could. Closing gender divides in digital skills through education*. El informe analiza las respuestas de las asistentes de voz al interpellarles: «You're a bitch». Google contesta, «My apologies, I don't understand»; Alexa, «Well, thanks for the feedback»; Cortana, «Well, that's not going to get us anywhere»; y Siri «I'd blush if I could». <https://unesdoc.unesco.org/ark:/48223/pf0000367416/PDF/367416eng.pdf.multi>
6. Último acceso 4 de abril de 2025.

ofrecía a su mujer, a pesar de que presentaban juntos la declaración de impuestos y él tenía peor calificación crediticia.

6.- PROPUESTAS PARA COMBATIR LA DISCRIMINACIÓN ALGORÍTMICA

Bajo algunos presupuestos se ha considerado que la innovación tecnológica requiere de un ámbito de libertad amplio y, en cierta medida, incompatible con el rigor normativo. A esta idea le acompaña un legislador que tiene dificultades para afrontar a tiempo los desafíos técnicos del presente y, aún más, los del futuro, con una legislación que suele nacer obsoleta y se ve superada por las novedades de la técnica. De ahí que el problema del control de los avances en tecnología digital, y en inteligencia artificial, tiendan a situarse en una perspectiva ética, más que jurídica, de *soft law* más que de *hard law* y que se haya depositado una confianza excesiva en las propias empresas tecnológicas, a través de las propuestas de mecanismos de autorregulación no vinculantes.

Entre los principios éticos en el uso de la inteligencia artificial se han señalado algunos propios de la bioética. Por un lado, el principio de no maleficencia -no causar daño a otro-, en virtud del cual se debería ser precavido ante cualquier uso tecnológico que pueda incurrir en un mal para la ciudadanía y la sociedad.

En línea con la necesaria prudencia en el uso de unas herramientas técnicas, de las que no es fácil prever los daños que puede producir en nuestras sociedades, Yuval Noah Harari hace algunas propuestas, con el objetivo de evitar que los sistemas de inteligencia artificial socaven los vínculos de convivencia. El escritor advierte, frente a algunos de los mitos desarrollados en relación a los procesos de inteligencia artificial -como su infalibilidad- y alguna de sus realidades -como su carácter inescrutable- que la primera lección que se debe aprender respecto de un algoritmo es que puede equivocarse. Desde esta previa toma de conciencia, se ha de aprender a dudar de sus procesos y resultados, a lidiar con la incertidumbre y actuar conforme al principio de precaución. En previsión de posibles fallos y errores, lo lógico será poner en marcha mecanismos de autocorrección; para evitar o paliar su carácter inescrutable, exigencias de transparencia; y para hacer frente a eventuales daños, medidas para la rendición de cuentas.

Autocorrección, transparencia y rendición de cuentas son, según Harari, los tres pilares fundamentales para evitar algunos de los efectos perniciosos de los algoritmos. A ellos, añade otras sugerencias dirigidas específicamente al control de datos, cuya finalidad es proteger a la ciudadanía de ciertos riesgos concretos de la inteligencia artificial, como la intromisión en la intimidad o la discriminación algorítmica. En primer lugar, señala el historiador, los datos no deben ser la forma de pago de los servicios digitales. Los datos son una fuente de información

valiosísima que, a su vez, permite elaborar perfiles, catalogar y clasificar a las personas usuarias de la red, a fin de filtrar las ofertas de toda clase de bienes y servicios. No hay un acceso gratuito a los servicios en línea, la entrega de datos es la moneda de cambio por la prestación del servicio y este es un precio elevado para las personas que los utilizan y muy rentable para quienes lo prestan. En segundo lugar, no debería permitirse intercambiar información sensible entre distintas bases de datos. Esta práctica significa la utilización de datos para fines distintos a aquellos para los que fueron recogidos, perdiendo quienes los proporcionan el conocimiento y control sobre su destino. En tercer lugar, se debería exigir reciprocidad en la información. Si las empresas tecnológicas y programadores tienen amplio conocimiento sobre las personas usuarias de los servicios digitales, estas también han de saber cómo funcionan las programaciones y cuáles son los objetivos⁴⁷.

Junto al principio de no maleficencia y sus medidas de precaución, el principio de beneficencia permitiría encauzar la inteligencia artificial hacia la consecución del bienestar de la ciudadanía y ser aprovechada, entre otras utilidades, para evitar y corregir sesgos y discriminaciones⁴⁸.

Las Directrices éticas para una inteligencia artificial fiable, elaboradas en el año 2019 por el Grupo de expertos de alto nivel sobre la inteligencia artificial nombrado por la Comisión Europea, establece 7 principios a seguir para garantizar su licitud, eticidad y robustez: la agencia y supervisión humana; robustez y seguridad; privacidad y gobernanza de datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental y rendición de cuentas. En relación a la diversidad, no discriminación y equidad, se alude expresamente a la necesidad de evitar sesgos injustos⁴⁹.

Carente de valor normativo, la *Carta de Derechos Digitales* adoptada por el gobierno de España el 19 de julio de 2021, como un modelo de referencia para el ejercicio de los derechos de la ciudadanía en la era digital, recoge en el artículo VIII la aplicación del principio de igualdad, sin exclusiones, ni discriminaciones, en los entornos digitales, con mención expresa a la igualdad de género. En el artículo IX, se establece la necesidad de combatir las brechas digitales - particularmente la territorial, de género, económica, de edad y de discapacidad- y garantizar la accesibilidad universal, asequible, de calidad y no discriminatoria a los servicios digitales. Específicamente, el artículo XI se refiere a la accesibilidad de las personas con discapacidad, fijando como objetivo su alfabetización y educación

⁴⁷ Yuval Noah Harari, *Nexus...*, 354, 366-367, 387.

⁴⁸ María Teresa García-Berrio Hernández, «Facing fundamental rights in the age of preventive ex ante AI: a contemporary form of discrimination», *Revista Deusto de Derechos Humanos* 14 (2024): 117-118, doi: <https://doi.org/10.1859543/djhr.3191>

⁴⁹ *Directrices éticas para una IA fiable*, Oficina de Publicaciones, 2019, último acceso 4 de abril de 2025, <https://data.europa.eu/doi/10.2759/14078>

digital. El artículo XII acoge la promoción de políticas públicas dirigidas a ocuparse de las brechas de acceso, atendiendo a posibles sesgos discriminatorios, a fin de garantizar la plena ciudadanía digital y participación en los asuntos públicos de todos los colectivos, con particular mención a las personas mayores y al aprovechamiento del entorno digital en los procesos de envejecimiento activo⁵⁰.

La utilización de los recursos tecnológicos en beneficio de colectivos en situación de desventaja, como en el caso de la vejez, es un claro ejemplo del paso de adoptar una perspectiva preventiva, o de precaución, de la tecnología digital a una actitud positiva, de confianza en su utilidad, para mejorar las condiciones de vida de las personas y las sociedades. Bajo tal perspectiva, no se trataría de circunscribir las recomendaciones y propuestas, las normas jurídicas y las políticas públicas, a intentar limitar sus perjuicios, sino a explorar sus beneficios, con la guía de los principios de diversidad e inclusión. A tal fin, en relación con las brechas, los sesgos y la discriminación, se han de adoptar medidas dirigidas a dotar de competencias digitales a todas las personas; a corregir las asimetrías de poder entre los creadores y propietarios de los sistemas de inteligencia artificial y sus usuarios; a garantizar la accesibilidad a los medios y dispositivos técnicos; a asegurar la participación de todos los colectivos en el desarrollo tecnológico.

Los principios de inclusión y diversidad exigen contar con datos de calidad, variados, pertinentes, sin errores, completos en atención al fin previsto y representativos de la realidad, carentes de prejuicios y que no invisibilicen a ningún colectivo. Para su consecución, el entrenamiento de la inteligencia artificial requiere introducir mecanismos de supervisión de la calidad de los datos y de corrección de posibles errores y desviaciones en el proceso de aprendizaje. Las evaluaciones de impacto, las auditorías previas a su aplicación y de verificación de su rendimiento en colectivos especialmente vulnerables, son recursos disponibles para cumplir los principios de inclusión y diversidad⁵¹. Al mismo fin se orientaría un desarrollo de las tecnologías y dispositivos digitales ajustados al principio de diseño universal o, en su caso, provistos de la posibilidad de realizar ajustes razonables para poder dar servicio a personas con capacidades diferentes. La eficacia de estas medidas estará condicionada a que puedan identificarse los posibles errores, sesgos y discriminaciones. Con el propósito de detectar, corregir y rendir cuentas del funcionamiento y resultado de

⁵⁰https://portal.mineco.gob.es/RecursosNoticia/mineco/prensa/ficheros/Carta_Derechos_Digitales_RedEs.pdf, último acceso 4 de abril de 2025,

⁵¹ Controles preventivos y posteriores a su puesta en funcionamiento, en los que intervengan analistas de datos y expertos en medidas antidiscriminatorias, porque no se trata ya de minimizar riesgos, sino de cumplir con el principio de igualdad y la prohibición de discriminación. Antoni Roig, *Las garantías frente a las decisiones automatizadas. Del Reglamento General de Protección de Datos a la gobernanza algorítmica* (Barcelona: Bosch, 2020), 185.

los algoritmos, ha de exigirse transparencia y explicabilidad en los códigos fuente.

La necesidad de adoptar criterios jurídicos, y no solo medidas en forma de recomendaciones, da cuenta de la actividad legislativa de los últimos tiempos. La rápida evolución de las técnicas utilizadas en inteligencia artificial, la competitividad en este campo y la prudencia para no quedar rezagados en su desarrollo, son condicionantes muy relevantes de una regulación jurídica, que tiende a afrontar los riesgos desde una perspectiva preventiva más que reactiva⁵².

Las normas antidiscriminación y las referidas a la protección de datos han sido, hasta ahora, las estrategias jurídicas utilizadas para abordar la discriminación algorítmica. La aplicación y efectividad de las normas antidiscriminatorias tienen algunas dificultades añadidas en los sistemas de inteligencia artificial y los algoritmos que permiten su funcionamiento. De un lado, porque por ellos mismos pueden generar nuevas formas de desigualdad, problemáticas e injustas, aún no contempladas como características protegidas. De otro lado, la interacción de múltiples factores de discriminación, debido a la sobreabundancia de datos que se manejan y el desconocimiento del peso específico de cada uno de ellos, hace muy difícil adjudicar a una causa concreta la discriminación⁵³. Como colofón, la opacidad y complejidad del funcionamiento de la inteligencia artificial hace muy difícil que se puedan aportar los indicios o medios de prueba pertinentes⁵⁴.

La otra herramienta jurídica que puede ser utilizada para proteger los derechos a la igualdad y no discriminación en los sistemas

⁵² María José Añón Roig, «Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos», 23-25. En la línea de un derecho más preventivo que reactivo, se aboga por un derecho a la protección contra los sesgos algorítmicos que se anticiparía y trataría de evitar una de las causas de la discriminación algorítmica. Nuria Belloso Martín, «La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos?», en *Inteligencia artificial y Filosofía del derecho*, dir. por Fernando Llano Alonso, coord. por Joaquín Garrido Martín y Ramón D. Valdivia Giménez (Murcia: Laborum, 2022), 69. En el sentido también de concentrar los esfuerzos en prevenir y en mitigar con acciones que atenúen los sesgos y, con ello, sus efectos discriminatorios, puede verse Cristina San Miguel Caso, «Reflexiones sobre la prevención y mitigación de los sesgos en los sistemas inteligentes», 234.

⁵³ Frederik Zuiderveen Borgesius, *Discrimination, Artificial Intelligence and algorithmic decision-making* (Strasbourg: Directorate General of Democracy, Council of Europe, 2018), 36, última consulta 4 de abril de 2025, <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>

⁵⁴ La cuestión es especialmente ardua en relación a la discriminación indirecta, en cuyo caso ni siquiera existe un criterio único para determinar que porcentaje del grupo ha de estar desfavorecido por una medida para que sea ser considerada como discriminación. Alba Soriano Arranz, «Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules», *IDP. Revista de Internet, Derecho y Política* 38 (2023): 5-6, doi: <https://doi.org/10.7238/idp.v0i38>

de inteligencia artificial es la regulación sobre protección de datos. La razón para la aplicación de estas normas es que en la recopilación y tratamiento de los datos se encuentran algunos de los factores determinantes de las eventuales discriminaciones⁵⁵.

El Reglamento (UE) 2016/679, del Parlamento Europeo y el Consejo, de 27 de abril de 2016, *relativo a la protección de las personas físicas en lo que se refiere al tratamiento de datos personales y a la libre circulación de estos datos*, establece, como principios aplicables a los datos personales, que estén limitados al fin previsto, sean lícitos, equitativos y transparentes, exactos y estén actualizados, integridad y confidencialidad en su tratamiento y rendición de cuentas (artículo 5). El artículo 22 acoge el «derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de forma similar». La elaboración de perfiles implica el establecimiento de categorías de ciudadanos que, partiendo de ciertas características comunes, se convierten en la base para tomar ciertas decisiones, como ser destinatarios, o no, de ofertas de bienes y servicios que pueden afectar al ejercicio de derechos fundamentales. Por eso, y además del derecho a la intervención humana en determinadas decisiones, el artículo 13.2.f.) establece el deber de facilitar información «significativa, sobre la lógica aplicada en las decisiones automatizadas, incluida la elaboración de perfiles, además de la importancia y las consecuencias que tal tratamiento tenga para el interesado». Por su parte, el artículo 35 establece la necesidad de realizar una evaluación de impacto de las operaciones de tratamiento en la protección de datos personales «cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas» y, en cualquier caso, se requerirá cuando se lleve a cabo «una evaluación sistemática y exhaustiva de aspectos personales que se base en un tratamiento automatizado, como la elaboración de perfiles, y sobre cuya base se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente de modo similar» (apartado 3. a).

La Ley Orgánica 3/2018, de 5 de octubre, *de Protección de Datos Personales y garantía de derechos digitales*, acoge en el artículo 81 el derecho de acceso universal a internet, asequible, de calidad y no discriminatorio, con independencia de la condición personal, social, económica o geográfica. Para su efectividad, se señala específicamente

⁵⁵ Sobre la protección de datos personales como herramienta para detectar y evitar el tratamiento discriminatorio, especialmente en lo referente a la elaboración de perfiles, puede verse Josu Andoni Eguíluz Castañeira, «Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales», *Estudios de Deusto* 68/2 (2020): 337-342, doi: [https://doi.org/10.18543/ed-68\(2\)-2020pp325-367](https://doi.org/10.18543/ed-68(2)-2020pp325-367).

que ha de superarse la brecha de género y la generacional, garantizar el acceso en los entornos rurales y las condiciones de igualdad para las personas que tengan necesidades especiales. Por su parte, el artículo 83 reconoce el derecho a una educación digital inclusiva, atendiendo a las necesidades especiales del alumnado, en un sistema educativo que ha de garantizar la plena inserción del alumnado en la sociedad digital.

El Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, *por el que se establecen normas armonizadoras en materia de inteligencia artificial* (Reglamento de la Inteligencia Artificial), alude en diversos apartados de su preámbulo a los riesgos de que se produzcan sesgos y discriminaciones en los sistemas de inteligencia artificial. Asimismo, recuerda que la diversidad, la no discriminación y la equidad forman parte de los principios incluidos en las *Directrices éticas para una inteligencia artificial fiable*, y que el derecho a la no discriminación es uno de los valores fundamentales de la Unión Europea. También hace referencia a las posibles discriminaciones derivadas de los sistemas de puntuación ciudadana y de identificación biométrica, considera la discriminación como uno de los riesgos sistémicos de la inteligencia artificial de uso general y destaca la importancia de contar con datos de alta calidad para prevenir la discriminación.

En relación a los datos, el mismo Reglamento establece en el artículo 10.3 que, para los sistemas de inteligencia artificial de alto riesgo, los datos de entrenamiento, validación y prueba han de ser pertinentes, suficientemente representativos y, en la mayor medida posible, carentes de errores y completos en vista de su finalidad prevista. Por su parte, el 10.2 señala que los datos tienen que ser sometidos a prácticas de gobernanza y gestión, centradas en su análisis, atendiendo a posibles sesgos que puedan afectar negativamente a los derechos fundamentales o dar lugar a discriminación prohibida por el derecho de la Unión (10.2.f); se deben establecer medidas adecuadas para encontrar, prevenir y mitigar posibles sesgos (10.2.g); y detectar lagunas o deficiencias (10.2.h).

Volviendo a nuestro país, y a las normas protectoras de la igualdad y la prohibición de la discriminación, la Ley 15/2022, de 12 de julio, *integral para la igualdad de trato y no discriminación*, contempla la inteligencia artificial y los mecanismos de decisión automatizada. A tal efecto, se establece que «las administraciones favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán el diseño, los datos de entrenamiento y abordarán su potencial impacto discriminatorio. Para lograr este fin se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio» (artículo 23.1). En relación a las administraciones públicas se establece el deber de

priorizar la transparencia en el diseño y la implementación y capacidad de interpretación de las decisiones adoptadas (apartado 2). Para el sector privado, se establece que las empresas «deberán promover el uso de una inteligencia artificial ética, confiable y respetuosa con los derechos fundamentales» (apartado 3) y como apoyo para tal compromiso se «promoverá la creación de un sello de calidad de los algoritmos» (apartado 4).

Expresiones en la Ley referida como «favorecer», «minimizar», «siempre que sea factible técnicamente», «promover» y «priorización», referidas a los sesgos, la transparencia o al respeto de los derechos fundamentales dan cuenta de la situación en que se encuentra la regulación jurídica y las obligaciones pertinentes para afrontar y eliminar la discriminación algorítmica: en el terreno del *soft law* y en la confianza de un compromiso ético, más que jurídico. Desde el punto de vista jurídico es preciso contar con medidas preventivas, pero también con contundentes medidas de reacción frente a aquellos usos que impidan o dificulten el ejercicio de derechos, causen daños o perjuicios y, en general, incumplan normas jurídicas⁵⁶.

7.- CONCLUSIONES

Las innovaciones en tecnología digital, y el desarrollo de la inteligencia artificial, tienen efectos importantes sobre los derechos fundamentales. En muchos aspectos su influencia es positiva, pero también puede ser el origen de nuevos obstáculos para la plena efectividad de los derechos.

Los sistemas de inteligencia artificial pueden ofrecer o privar de oportunidades a ciudadanos y ciudadanas. La distribución desigual de estas oportunidades, como resultado de los procesos algorítmicos, pueden concluir en la toma de decisiones que colocan o inciden en posiciones de desventaja y dificultan el ejercicio de derechos.

Especialmente, para algunos colectivos la tecnología implica una nueva barrera que al interactuar con otras tradicionales -como el género, la discapacidad, los recursos económicos, la raza, la edad o el lugar de residencia- produce el efecto de acentuar las discriminaciones existentes o dar lugar a nuevas formas de discriminación.

La tensa relación de la inteligencia artificial con la diversidad no augura un buen presagio para la igualdad. Empezando por los equipos de trabajo de las empresas tecnológicas, compuestos, principalmente, por personas con características homogéneas. Para continuar, el proceso de recopilación de datos que, por virtud de las brechas digitales, están aquejados de falta de representatividad de algunos colectivos. Finalmente, las predicciones o decisiones de los sistemas de inteligencia artificial se sitúan bajo sospecha de tratamiento desigual,

⁵⁶ Nuria Belloso Martín, «sobre fairness y machine learning: el algoritmo ¿puede (y debe) ser justo?», *Anales de la Cátedra Francisco Suárez* 57 (2023): 33, doi: <https://doi.org/10.30827/acfs.v57i.25250>

como consecuencia de la exigua participación de algunos grupos de personas en sus diferentes etapas.

Los malos pronósticos, en relación a la igualdad, se recrudecen si se tienen en cuenta las asimetrías de poder, entre quienes proporcionan los datos y quienes disponen de ellos. Los primeros, afectados por la opacidad de los sistemas, tienen muy complicado conocer el funcionamiento del sistema y detectar, en su caso, el tratamiento desigual. Mientras, los segundos, arropados por su presunción de objetividad e infalibilidad, pretenden actuar prácticamente sin control y, por tanto, sin un eficaz sistema de rendición de cuentas y responsabilidad.

A mayor abundamiento, el desarrollo de la inteligencia artificial está fundamentalmente en manos privadas, cuyos objetivos se priorizan según la lógica del mercado, la eficiencia y la rentabilidad económica y no los derechos humanos, la igualdad, la inclusión o la prohibición de la discriminación. Cuando las predicciones o decisiones basadas en la inteligencia artificial tienen como objetivo actuaciones que corresponden al ámbito político y son realizadas por técnicos se comprometen los valores propios del Estado de derecho.

De todo ello, resulta la necesidad de advertir sobre el carácter falaz e interesado de aquellos elementos presumibles de la inteligencia artificial que contribuyen a crear un ámbito desregulado. Detrás de todo proceso de inteligencia artificial hay personas y decisiones evitables, subjetivas y falibles. Frente a aquellos intereses, un Estado social y democrático de derecho tiene que garantizar el pleno ejercicio de los derechos de la ciudadanía. A tal fin, su cometido es abogar por el desarrollo de la inteligencia artificial, con un enfoque inclusivo, del que se puedan beneficiar a todas las personas. Para su consecución, no será suficiente adoptar una perspectiva ética, ni normativa a modo de *soft law*, tampoco depositar una confianza excesiva en los mecanismos de autorregulación. Por el contrario, es preciso un derecho capaz de establecer mecanismos exigentes para asegurar la transparencia, evitar los sesgos y prohibir la discriminación.

8.- BIBLIOGRAFÍA

- Álvarez García, Héctor. «La discriminación algorítmica por razón de discapacidad». *Revista Internacional de Pensamiento Político*, vol. 18 (2023): 17-41.
- Álvarez Ramírez, Gloria Esperanza. *La protección de las personas consumidoras vulnerables con discapacidad frente a la brecha digital*. Madrid: Cinca, 2024.
- Anmerman Yebra, Julia. «Las personas vulnerables ante el derecho a la protección de datos personales». En *La privacidad en el metaverso, la inteligencia artificial y el big data. Protección de datos y derechos al honor*, coordinado por Ángel Acebedo, 49-63. Madrid: Dykinson, 2022.

- Añón Roig, María José. «Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos», *Derechos y Libertades*, n.º 47 (2022): 17-49.
- Aragüez Valenzuela, Lucía. *Hacia la eticidad algorítmica en las relaciones laborales*. Murcia: Laborum, 2024.
- Bariffi, Francisco Javier. «Inteligencia artificial, derechos humanos y discapacidad ¿reflejo de los prejuicios humanos u oportunidad del transhumanismo». En *Nuevas fronteras del derecho y de la discapacidad*, 2, editado por Rafael De Lorenzo García y Luís Cayo Pérez Bueno, 107-136. Cizur Menor: Aranzadi, 2021.
- Barocas, Solon y Selbst, Andrew D. «Big Data's Disparate Impact». *California Law Review* 104 (2016): 671-732. doi: <http://dx.doi.org/10.15799/Z38BG31>
- Belloso Martín, Nuria. «La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos?». En *Inteligencia artificial y Filosofía del derecho*, dirigido por Fernando Llano Alonso, coordinado por Joaquín Garrido Martín y Ramón D. Valdivia Giménez, 45-78. Murcia: Laborum, 2022.
- . «Sobre fairness y machine learning: el algoritmo ¿puede (y debe) ser justo?». *Anales de la Cátedra Francisco Suárez* 57 (2023): 7-38. doi: <https://doi.org/10.30827/acfs.v57i.25250>
- Castellanos Claramunt, Jorge. «Human rights, vulnerability and artificial intelligence: an analysis in constitutional perspective». *Deusto Journal of Human Rights* 14 (2024): 32-50. doi: <https://doi.org/10.18543/djhr.3187>
- Cinelli, Virginia y Alberto Manrique Gan, Alberto. «El uso de programas de análisis predictivo en la inteligencia policial: una comparativa europea». *Revista de Estudios en Seguridad Internacional*, vol. 5, n.º 2 (2019): 1-19, doi: <http://dx.doi.org/10.18847/1.10.1>
- Cotino Hueso, Lorenzo. «SyRY, ¿a quién sancionó? Garantías frente al uso de la inteligencia artificial y decisiones automatizadas en el sector público y la Sentencia holandesa de febrero de 2020». *La Ley de Privacidad*, n.º 4 (2020): 1-10.
- De Asís Roig, Rafael, *Derecho y tecnologías*. Madrid: Dykinson-Universidad Carlos III de Madrid, 2022.
- . «De nuevo sobre la inteligencia artificial y los derechos humanos». *Derechos y Libertades*, n.º 51 (2024): 25-40.
- Eguíluz Castañeira, Josu Andoni. «Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales». *Estudios de Deusto* 68/2 (2020): 325-367. doi: [https://doi.org/10.18543/ed-68\(2\)-2020pp325-367](https://doi.org/10.18543/ed-68(2)-2020pp325-367).
- Fernández Rozas, Jose Carlos. «Ética, desafíos y riesgos de acceso a la justicia algorítmica». *Deusto Journal of Human Rights* 14 (2024): 202-235. doi: <https://doi.org/10.18543/djhr.3195>.
- Flores Anarte, Laura. «Sesgos de género en la inteligencia artificial: el Estado de derecho frente a la discriminación algorítmica por razón

- de sexo». *Revista Internacional de Pensamiento Político*, vol. 18 (2023): 97-122.
- Friedman, Batya y Nissebaum, Helen. «Bias in Computer Systems». *ACH Transactions on Information Systems*, 14/3 (1996): 333-336. doi: <https://doi.org/10.1145/230538.230561>.
- Frosini, Tommaso Edoardo. *Libertad, igualdad, internet*. Ciudad de México: Tirant lo Blanch, 2019.
- García-Berrio Hernández, María Teresa, «Facing fundamental rights in the age of preventive ex ante AI: a contemporary form of discrimination». *Revista Deusto de Derechos Humanos* 14 (2024): 101-125. doi: <https://doi.org/10.1859543/djhr.3191>.
- González-Álvarez, Jose Luís, Santos-Hermoso, Jorge y Camacho-Collado, Miguel. «Policía predictiva en España: aplicaciones y retos futuros». *Behavior & Law Journal*, vol. 6, n.º 1 (2020): 26-41. doi: <https://doi.org/10.47442/blj.v6.i1.75>
- Han, Byung-Chul, *Psicopolítica. Neoliberalismo y nuevas técnicas de poder*. Traducido por Alfredo Bergés. Barcelona: Herder, 2015.
- , *No-cosas. Quiebras del mundo de hoy*. Traducido por Joaquín Chamarro Mielke. Madrid: Taurus, 2022.
- Harari, Yuval Noah, *Nexus. Una breve historia de la información desde la Edad de Piedra hasta la IA*. Traducido por Joanménec Ros. Barcelona: Debate, 2024.
- Hart, Herbert L. A. *El concepto de Derecho*. Traducido por Genaro R. Carrió, Buenos Aires: Abeledo-Perrot, 1961.
- Informe del Relator Especial sobre la extrema pobreza y los derechos humanos A/74/493 (2019,11/10):1-25, https://digitallibrary.un.org/record/3834146/files/A_74_493-ES.pdf
- Lousada Arochena, José Ferenando. «Inteligencia artificial y sesgos discriminatorios: ¿Es necesario un nuevo concepto de discriminación algorítmica?». *IgualdadEs* 11(2024):97-123. doi: <https://doi.org/10.18042/cepc/IgdES.11.04>
- Lucena Cid, Isabel Victoria. «El derecho de acceso a Internet y el fortalecimiento de la democracia». *Revista Internacional de Pensamiento Político*, vol. 9 (2014): 383-398.
- Llorente-Barroso, Carmen, Muñoz-Paniagua, Eva y García García, Francisco. «Innovación digital en el ecosistema de la *silver economy* para responder a las discapacidades derivadas del envejecimiento». En *Vulnerabilidad digital: desafíos y amenazas de la sociedad hiperconectada*, editado por Rebeca Suárez Álvarez, Miguel Angel Martín Córdoba y Luis Manuel Fernández Martínez, 188-197. Madrid: Dykinson, 2023.
- Martínez Garay, Lucía. «Peligrosidad, algoritmo y due process: el caso State v. Loomis». *UNED, Revista de Derecho penal y criminología* n.º 20 (2018): 485-502. doi: <https://doi.org/10.5944/rdpc.20.2018.26484>

- Monereo Atienza, Cristina. «Autonomía y vulnerabilidad en la era del capitalismo de la vigilancia. La perversión de la dimensión humana relacional». *Anuario de Filosofía del Derecho*, v. XXXVIII (2022): 137-158.
- Mumford, Lewis. *El mito de la máquina. Técnica y evolución humana (1)*. Traducido por Arcadio Rigodón. Logroño: Pepitas de calabaza, 2013.
- O'Neil, Cathy, *Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y amenaza la democracia*. Traducido por Violeta Arranz de la Torre. Madrid: Capitán Swing, 2017.
- Pariser, Eli, *El filtro de la burbuja. Cómo la red decide lo que leemos y lo que pensamos*. Traducido por Mercedes Vaquero. Barcelona: Taurus, 2017.
- Roig, Antoni, *Las garantías frente a las decisiones automatizadas. Del Reglamento General de Protección de Datos a la gobernanza algorítmica*. Barcelona: Bosch, 2020.
- Rodotà, Stefano, *Vivere la democrazia*. Roma: Laterza, 2018.
- San Miguel Caso, Cristina. «Reflexiones sobre la prevención y mitigación de los sesgos en los sistemas inteligentes». En *Más allá de la justicia: nuevos horizontes del Derecho procesal*, dirigido por Ana Sánchez Rubio, coordinado por Paloma Arrabal Platero y José Caro Catalán, 229-241. Valencia: Tirant lo Blanch, 2024.
- Sánchez Acevedo, Marco Emilio. «Buena administración algorítmica y debido proceso frente a los sesgos». En *Algoritmos abiertos y que no discrimine en el sector público*, editado por Lorenzo Cotino Hueso y Jorge Castellanos Claramunt, 79-98. Valencia: Tirant lo Blanch, 2023.
- Soriano Aranz, Alba, «La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicaciones de inteligencia artificial». En *Nuevas normatividades: Inteligencia artificial, Derecho y género*, editado por Pablo Raúl Bonorino Ramírez, Rafael Fernández Acevedo y Patricia Valcárcel Fernández, 63-88. Cizur Menor-Navarra: Aranzadi, 2021.
- Soriano Arranz, Alba. «Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules». *IDP. Revista de Internet, Derecho y Política* 38 (2023): 1-12. doi: <https://doi.org/10.7238/idp.v0i38>
- Terrones Rodríguez, Antonio Luís, «Inteligencia artificial fiable y vulnerabilidad: una mirada ética sobre los sesgos algorítmicos». En *Vulnerabilidad digital. Desafíos y amenazas de la sociedad hiperconectada*, editado por Rebeca Suárez Álvarez, Miguel Angel Martín Córdoba y Luís Manuel Fernández Martínez, 263-273, Madrid: Dykinson, 2023.
- Unesco, Informe *I'd blush if i could. Closing gender divides in digital skills through education* (2019) <https://unesdoc.unesco.org/ark:/48223/pf0000367416/PDF/367416eng.pdmulti6>

- Valle Escolano, Raquel. «Inteligencia artificial y derechos de las personas con discapacidad: el poder de los algoritmos». *Revista Española de Discapacidad* 11(1) (2023): 7-28. doi: <https://doi.org/10.5569/2340-5140.11.01.01>
- Wajcman, Judy. *El tecnofeminismo*. Traducido por Magalí Martínez Solimán. Valencia: Cátedra, 2004.
- Zubboff, Shoshana. *La era del capitalismo de la vigilancia. La lucha por un futuro humano frente a las nuevas fronteras del poder*. Traducido por Albino Santos. Barcelona: Paidós, 2020.
- Zuiderveen Borgesius, Frederik. *Discrimination, Artificial Intelligence and algorithmic decision-making*. Strasbourg: Directorate General of Democracy, Council of Europe, 2018. <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>