

### 3.2 Algunas consideraciones éticas y jurídicas a propósito del desarrollo y uso de los sistemas de inteligencia artificial

*Ana Aba Catoira  
Profesora de Derecho Constitucional  
Universidad de A Coruña (España)*

La revolución tecnológica verdaderamente disruptiva nos sitúa ante continuos desafíos presentes y futuros. En este escenario revolucionado la Inteligencia Artificial (IA) representa el desafío número uno, al tener entre sus propósitos el desarrollo de sistemas inteligentes que imiten al ser humano<sup>136</sup>.

Lógicamente este entorno social y tecnológico nos beneficia en muchos sentidos al simplificar muchos trámites y procedimientos a través de dispositivos y aplicaciones y al ofrecernos unos servicios antes desconocidos o inalcanzables para muchos de nosotros. Además, no solo beneficios individuales o colectivos sino también la innovación y desarrollo al servicio del interés general como los avances médicos o farmacológicos o la seguridad nacional o ciberseguridad entre otros.

No obstante, la gran preocupación global se centra en el buen uso de la tecnología, pues lo contrario, un uso opaco y discriminatorio, constituye un grave riesgo para los derechos fundamentales y el buen funcionamiento de los sistemas democráticos. En palabras de Barrio, las tecnologías digitales emergentes ofrecen diversas oportunidades para mejorar la vida de los ciudadanos, pero es importante centrarse en el uso que se les da, ya que esto puede determinar si son un medio para reforzar los derechos y libertades existentes o si pueden poner en riesgo algunos de ellos, como el derecho a la privacidad y protección de datos, el desarrollo personal o la seguridad de los menores en Internet<sup>137</sup>.

Y ciertamente no se trata, de ningún modo, de un problema menor, pues hablamos de vulneración –masiva- de derechos fundamentales siendo algunos directa e intensamente afectados como ocurre con el derecho a la privacidad y protección de datos personales y el derecho efectivo a la igualdad de trato y no discriminación. Obviamente, riesgos de daño a causa de impacto del big data y del uso de algoritmos para la toma de decisiones que afectan a personas.

Por todo lo expuesto, resulta conveniente conocer y evaluar las implicaciones ético-jurídicas que puede tener el avance de la IA y otras tecnologías digitales en el ámbito de los derechos y libertades y las obligaciones de las personas. Tanto es así que resulta necesario establecer una adecuada regulación con un fundamento ético sólido<sup>138</sup>.

<sup>136</sup> El término "inteligencia artificial" (artificial intelligence) fue acuñado por John McCarthy en 1956 durante la Conferencia de Dartmouth, un evento histórico que reunió a algunos de los mejores científicos de la época para discutir la posibilidad de crear una máquina que pudiera pensar como un ser humano. Sin embargo, los conceptos e ideas que estaban detrás de la inteligencia artificial se remontan a mucho antes.

<sup>137</sup> M. Barrio Andrés, *Formación y evolución de los derechos digitales*, Ediciones Jurídicas Olejnik, Colección Biblioteca de Derecho Civil, Santiago de Chile, 2021.

<sup>138</sup> Aba Catoira, Ana: «La garantía de los derechos como respuesta frente a los retos tecnológicos» en F. Balaguer Callejón y L. Cotino Hueso (coords.) *Derecho Público de la Inteligencia Artificial*, Fundación Manuel Giménez Abad, Zaragoza, 2023, pp. 57-84; Castellanos Claramunt, Jorge: «Derecho e inteligencia artificial desde una perspectiva ética: estudio de las garantías y los límites legales de los derechos fundamentales en la era digital», en L. Fontestad Portalés y M.N. Jiménez López (dirs.), P.R. Suárez Xavier, L.A. Flórez Álvarez, y B. Andrés

En este sentido, las respuestas desde el Derecho deben tener unas sólidas bases éticas y, en esta línea, desde hace años, ya décadas, se viene estableciendo un marco ético para el desarrollo e implementación de las tecnologías. Se presenta como imprescindible la adopción de un enfoque de derechos humanos ante un entorno tecnológico difícilmente regulable que permita afirmar, de forma rotunda, que los derechos son el marco de referencia a la hora de abordar las tecnologías. De la aplicación de este enfoque se desprenden requisitos y exigencias éticas que han de cumplir los sistemas de inteligencia artificial para prevenir daños o efectos negativos como, por ejemplo, las decisiones prejuiciosas o estereotipadas.

En términos de exigencias éticas recurrimos al principio de precaución a día de hoy reconvertido en un principio jurídico al tener como contenido primordial la dignidad humana sobre el avance científico o tecnológico, en el sentido expresado en el artículo 10.1º de la Constitución española o en los convenios internacionales. Como expresión de este principio derivado de la dignidad y de los derechos encontramos las normas en materia de bioética o sobre responsabilidad de los daños ocasionados por las máquinas (robots).

Y ello sin olvidar la importancia que juegan las exigencias de transparencia y confiabilidad o el respeto del principio de igualdad de género que fundamenta el derecho de acceso equitativo muy presente en América Latina.

Así, en su discurso de 17 de octubre de 2019, Michelle Bachelet, Alta Comisionada de las Naciones Unidas para los Derechos Humanos, en su discurso “Derechos humanos en la era digital. ¿Pueden marcar la diferencia?” afirmaba: “Es esencial que en esta era digital prestemos especial atención a los derechos humanos... La revolución digital plantea un considerable problema de derechos humanos a escala mundial. Sus beneficios indudables no anulan sus riesgos evidentes”. Para después formular una pregunta: “¿Y abordamos estos desafíos mediante la ética o mediante los derechos humanos?”.

La respuesta que parece más adecuada y, en cuya línea se trabaja, es la ofrecida por la propia Bachelet que afirmaba que “los códigos éticos y el cumplimiento voluntario no constituyen, por sí mismos, una respuesta suficientemente enérgica para la escala del problema que afrontamos... No hay ningún segmento de la revolución digital que no pueda y no deba examinarse desde una perspectiva de derechos humanos”<sup>139</sup>.

Veamos entonces como se responde jurídicamente a estos problemas dimanados del uso de las tecnologías.

Por los que respecta al ámbito internacional, la Conferencia General de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) en el preámbulo del Proyecto de texto de la Recomendación sobre la ética de la inteligencia artificial de 2021, aborda la ética de la IA como una reflexión normativa sistemática, basada en un marco integral y evolutivo de valores, principios y acciones interdependientes, que puede guiar a las sociedades a la hora de afrontar de manera responsable los efectos conocidos y desconocidos de las tecnologías de la IA en los seres humanos, las sociedades y el medio ambiente y los ecosistemas, y les ofrece una base para aceptar o rechazar las tecnologías de la IA.

---

Segovia (coords.) *Justicia, proceso y tutela judicial efectiva en la sociedad post pandemia*, Aranzadi Thompson Reuters, 2022, pp. 69-84.

<sup>139</sup> En este sentido, Aba Catoira, Ana: «La era de la ciudadanía conectada digitalización y retos del futuro desde una perspectiva de género» en J. Castellanos Claramunt *Un estudio sobre el Estado autonómico: propuestas de mejora para el tercer decenio del Siglo XXI*, Dykinson, 2023, pp. 165-190.

La Unesco considera la ética como una base dinámica para la evaluación y la orientación normativas de las tecnologías de la IA, tomando como referencia la dignidad humana, el bienestar y la prevención de daños sin pretender su equiparación con el derecho, los derechos humanos ni verla como complemento normativo de las tecnologías.

En cuanto a la actividad del Consejo de Europa lo primero destacable es su liderazgo en la regulación ética de la IA a través de la adopción de distintas declaraciones y documentos entre los que destacan las Directrices sobre Inteligencia Artificial y Protección de Datos, aprobadas en enero de 2019, que promueven la protección de la dignidad humana y la salvaguarda de los derechos humanos y las libertades fundamentales en cualquier aplicación de IA con consecuencias para las personas y la sociedad, especialmente en los procesos de toma de decisiones. Asimismo, en la Declaración del Comité de Ministros sobre las capacidades manipuladoras de los procesos algorítmicos de febrero de 2019, el Consejo de Europa advierte sobre los efectos negativos que los procesos algorítmicos pueden tener en la autonomía cognitiva de los individuos y su derecho a formar opiniones y tomar decisiones independientes, lo que no solo debilita el ejercicio de los derechos humanos individuales, sino que puede poner en peligro los fundamentos mismos del Consejo de Europa.

Ahora bien, resulta necesario reflexionar sobre la adecuación de los marcos jurídicos y éticos existentes para hacer frente a las problemáticas derivadas de las tecnologías digitales y, especialmente, de la IA. En este orden de cosas, la línea de trabajo de la UE parece encaminarse a una tecnología “buena” por fiable y ética, basada en valores y en el respeto de los derechos fundamentales de la Unión consagrados en la Carta de los Derechos Fundamentales de la UE.

La Comisión Europea, en su Comunicado del 8 de abril de 2019, establece siete requisitos fundamentales para asegurar la fiabilidad de la inteligencia artificial entre los que se encuentran: la necesidad de intervención y supervisión humanas para garantizar que los sistemas de IA promuevan sociedades justas; la robustez y seguridad de los algoritmos para solucionar errores o inconsistencias que puedan surgir en todas las etapas del ciclo de vida de los sistemas de IA; la privacidad y la gestión adecuada de los datos personales; la transparencia, en el sentido de que los sistemas de IA deben ser trazables; la diversidad, la no discriminación y la accesibilidad de los sistemas de IA para adaptarse a las capacidades, competencias y necesidades humanas; la promoción del bienestar social y ambiental mediante sistemas de IA que fomenten la sostenibilidad y la responsabilidad ecológica, y, por último, la rendición de cuentas mediante la implementación de mecanismos que permitan controlar los resultados de los sistemas de IA.

En la actualidad, la principal normativa de la Unión Europea sobre inteligencia artificial es el proyecto de Reglamento de IA, que fue presentado por la Comisión Europea en abril de 2021. El Reglamento es una propuesta de regulación de la Unión Europea que busca establecer un marco de referencia para la producción, comercialización y uso de la inteligencia artificial en el territorio de la Unión. Este Reglamento establece una serie de requisitos para los sistemas de inteligencia artificial que sean considerados de alto riesgo, que incluyen la transparencia, la trazabilidad, la calidad de los datos y la evaluación de riesgos y daños. Igualmente, el Reglamento establece la creación de un registro de sistemas de IA de alto riesgo y una serie de obligaciones para los fabricantes y usuarios de dichos sistemas. Asimismo, también establece una serie de sanciones para los incumplimientos, que incluyen multas de hasta el 6% de la facturación anual de la empresa.

De este modo la Comisión Europea ha propuesto una regulación para la inteligencia artificial que establece cuatro niveles de riesgo y establece que todas las regulaciones deben ser aplicadas de manera igualitaria en todos los Estados Miembros, excluyendo el uso de la IA en el ámbito militar. En el nivel más alto de riesgo se encuentra lo que se considera «riesgo inaceptable», lo cual incluye sistemas de IA que se consideran una amenaza para la seguridad, los medios de vida y los derechos de las personas.

Los sistemas de IA que caen dentro de esta categoría serán prohibidos. A este respecto la Comisión ha dado algunos ejemplos de lo que considera un riesgo inaceptable, incluyendo un sistema de IA que manipule el comportamiento humano y fomente la violencia, como un juguete con asistencia de voz que pueda incitar a los menores a un comportamiento peligroso. También se incluye un sistema de «puntuación social» por parte de los gobiernos para diferenciar a los ciudadanos. En otro punto de «alto riesgo», se incluyen los usos de la inteligencia artificial en infraestructuras críticas que puedan tener un impacto en la salud de los ciudadanos, así como en la educación en caso de que permita hacer trampas en exámenes.

También se incluyen los componentes de la IA en cirugía, sistemas de contratación de personal, servicios públicos, legislación, inmigración y en la administración y justicia. Como no podía ser de otra manera, en todos estos campos, la inteligencia artificial estará sujeta a rigurosas obligaciones, que incluyen un análisis exhaustivo de riesgos, la capacidad de rastrear los resultados obtenidos, una documentación detallada, la supervisión humana y un alto nivel de resistencia a fallos.

En un nivel de riesgo más bajo, se incluyen sistemas como los chatbots, los cuales deben cumplir con un mínimo nivel de transparencia y los usuarios deben ser informados de que están interactuando con una máquina. Y el riesgo mínimo incluye todos los demás usos de la inteligencia artificial, como los videojuegos, las aplicaciones de imagen y otros sistemas que no implican riesgos. En estos casos, la nueva regulación no detalla medidas específicas a seguir. Otra cuestión significativa es que todos los sistemas de «identificación biométrica remota» se considerarán de alto riesgo. Así, la Comisión Europea no ha decidido prohibir directamente los sistemas de reconocimiento facial, aunque sí establece que se aplicarán requisitos estrictos como los mencionados anteriormente.

De modo que, en principio, el reconocimiento facial tendrá prohibido su uso en áreas públicas y en tiempo real para la aplicación de la ley. Esto significa que las autoridades no podrán utilizar el reconocimiento facial para prevenir posibles delitos, aunque existen algunas excepciones específicas, como buscar a un niño desaparecido, prevenir una amenaza terrorista inminente o detectar, localizar, identificar o enjuiciar a un sospechoso o perpetrador de un delito grave. Estos usos específicos estarán sujetos a la autorización de un órgano judicial u otro organismo independiente, y estarán limitados en cuanto a tiempo, ámbito geográfico y bases de datos buscadas.

En general, el Reglamento busca garantizar el respeto a los derechos fundamentales y los principios de transparencia, responsabilidad y seguridad en el uso de la inteligencia artificial. La propuesta de Reglamento se integra en un conjunto de medidas que pretenden abordar la problemática derivada del desarrollo y usos de sistemas de inteligencia artificial, como el “Libro Blanco sobre sobre la inteligencia artificial, un enfoque europeo orientado a la excelencia y la confianza”<sup>140</sup>, además de que estas medidas deben ser coherentes y

---

<sup>140</sup> Bruselas, 19.2.2020, COM (2020) 65 final.

complementarse con otras ya adoptadas por la Comisión en orden a prevenir y solucionar esta compleja problemática.

A mayor abundamiento se alinea con la estrategia digital general de la Comisión, que tiene como objetivo promover una tecnología al servicio de las personas y establecer un marco coherente, efectivo y proporcionado para garantizar que la IA se desarrolle de manera que respete los derechos de las personas y gane su confianza. Obviamente, absolutamente imprescindible, para cumplir los objetivos de la Década Digital.

Cabe mencionar también otras iniciativas europeas como la Estrategia de Datos de la UE, la Ley de Gobernanza de Datos y la Directiva relativa a los datos abiertos que establecen mecanismos y servicios de confianza para reutilizar, compartir y poner en común datos esenciales para el desarrollo de modelos de IA de gran calidad basados en datos. Lo anterior sin desconocer la existencia de otras normas como el Reglamento General de Protección de Datos (GDPR) y la Directiva de Servicios de la Sociedad de la Información (DSSI), que establecen requisitos para la protección de datos personales y la transparencia de los servicios en línea, respectivamente.

Retomando los derechos humanos a partir del impacto de las tecnologías digitales se debaten cuestiones que afectan al actual marco de los derechos humanos en cuanto si es suficiente o es necesario reconocer nuevos derechos; y, por otro, si, además de los derechos, es necesario utilizar otras herramientas.

Así, se discute sobre el necesario reconocimiento de nuevos derechos como derechos digitales o si, por el contrario, resultan suficientes los derechos ya reconocidos. En este orden de cosas, los defensores del reconocimiento de nuevos derechos afirman que el discurso actual de los derechos en el caso de que consideremos que se trata de propuestas que deben ser tenidas en cuenta, habrá que determinar un catálogo fundamentado y con un sistema de garantías que los haga eficaces. Según Barrio, se deberían reconocer nuevos derechos fundamentales stricto sensu en las Constituciones, como el acceso universal a Internet y la ciberseguridad, así como un derecho fundamental a la verdad para combatir las noticias falsas (fake news) y un derecho a la conciliación familiar y laboral no es suficiente para protegernos de las amenazas de las tecnologías convergentes.

Ahora bien, en esta discusión habrá de tenerse en consideración una cuestión vinculada al reconocimiento nuevos derechos y que se conoce como el problema de la inflación (cuantos más derechos reconocidos menos protección especial). En este plano también se aduce que la ampliación de un catálogo de derechos genera indeterminación y problemas de fundamentación, además, tal como se ha señalado, de un riesgo de pérdida de eficacia por un posible solapamiento con derechos ya reconocidos y garantizados en nuestras constituciones. Esto es, si los bienes jurídicos que se pretenden garantizar con el reconocimiento de nuevos derechos que responden al entorno digital, ya están suficientemente cubiertos con los derechos consagrados a nivel nacional e internacional.

En consonancia con la necesaria concreción de los derechos ante el paso del tiempo y las nuevas circunstancias y necesidades, así como amenazas y riesgos, inherentes a una concreta realidad social, los derechos deben ser redefinidos, actualizados, puestos al día, para no convertirse en categorías petrificadas o ineficaces. Ello significaría siguiendo a Pérez Luño que existe una

necesidad innegable de establecer un marco jurídico en torno al uso de las nuevas tecnologías NBIC y el desarrollo de la IA y la robótica avanzada<sup>141</sup>.

El artículo 18.4 de la Constitución española, inspirándose en el art. 35 de la Constitución portuguesa de 1976, supuso una gran innovación constitucional al limitar el uso de la informática para garantizar el honor y la intimidad de los ciudadanos. Esta cláusula constitucional estableció la base para el desarrollo normativo y jurisprudencial de un nuevo derecho fundamental autónomo, el derecho a la protección de datos personales. Ahora bien, con el derecho a la protección de datos personales no se ofrece protección suficiente a las personas en el entorno digital siendo necesario explorar y establecer un arco de garantías efectivas de los derechos de la ciudadanía en la era digital.

La Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales y la Carta de Derechos Digitales de 2021 que carece de fuerza normativa, pero que funciona como referencia para una futura ley reguladora de los derechos digitales<sup>142</sup>, persiguen este objetivo. El fundamento de la Carta, según se recoge en sus consideraciones previas, es «La dignidad de la persona, los derechos inalienables que le corresponden, el desarrollo libre de la personalidad, el respeto a la ley y a los derechos de los demás, son la base del orden político y la paz social». De modo que recoge un evidente enfoque constitucional que plasma una concepción de la persona y del Estado que es relevante tanto para el presente como para el futuro. El rápido progreso de la investigación científica, la invención y las tecnologías digitales o basadas en la digitalización, plantean la necesidad de garantizar que el marco normativo proteja los derechos individuales y colectivos de las personas, así como los valores constitucionales que son el fundamento de la convivencia.

En cualquier caso, un avance en materia legislativa y una guía o referencia para futuros desarrollos acordes a la realidad digital.

En cuanto a la discusión sobre la necesidad de establecer nuevas herramientas la Oficina del Alto Comisionado de Naciones Unidas para los Derechos Humanos, en su informe sobre “El derecho a la privacidad en la era digital” (2021), señalaba que los Estados deben establecer mecanismos de supervisión y reparación relacionados con la privacidad. La UNESCO, a través de su Comité Internacional de Bioética, emitió un informe el 15 de diciembre de 2021 sobre las Cuestiones Éticas de la Neurotecnología, en el que se encuentran también una serie de sugerencias generales y otras dirigidas a la propia UNESCO, a los Estados, a los investigadores, a las industrias, a los medios de comunicación e incluso al público en general. Entre las generales se señala: “a) Agregar protocolos a los tratados internacionales, como la Declaración Universal de los Derechos Humanos, para abordar los desafíos que plantean las neurotecnologías. b) Reforzar la Declaración Universal de los Derechos Humanos, considerando que la neurotecnología desafía los derechos humanos existentes y que se requerirán nuevas garantías en función de las posibilidades de vulneración. c) Elaborar una Nueva Declaración Universal de Derechos Humanos y Neurotecnología”.

Las recomendaciones a los Estados se resumen en la “concesión de un estatus positivo a los neuroderechos”, velando por que “sus leyes fundamentales reconozcan y garanticen claramente la integridad física y psíquica que permita a las personas el pleno goce de su identidad personal,

<sup>141</sup> PÉREZ LUÑO, Antonio E.: «Concepto y concepción de los derechos humanos (Acotaciones a la ponencia de Francisco Laporta)» en *Doxa: Cuadernos de Filosofía del Derecho*, Nº 4, 1987, pp. 47-66.

<sup>142</sup> Un análisis exhaustivo de la Carta en L. Cotino Hueso (coord.), *La Carta de Derechos Digitales*, Tirant lo Blanch, Valencia, 2022.

y el derecho a obrar de manera autónoma, y que sólo la ley pueda establecer los requisitos para limitarlo”. Y en relación con el público se recomienda:

“a) Enfatizar que cada individuo es el propietario de los datos que se recopilan de él o ella, y que solo pueden ser utilizados, publicados o comercializados en circunstancias excepcionales y solo con el consentimiento informado explícito. b) Tomar conciencia de los beneficios y riesgos potenciales de la neurotecnología, especialmente cuando afecta la integridad individual, influye en la percepción o induce a la toma de decisiones, y participa en debates públicos y otras acciones para examinar posibles abusos. c) Involucrarse en temas de neuroética y neuroderechos, de forma individual o mediante la formación de grupos de interés. d) Usar medios legales, incluyendo leyes y presión pública, para prevenir el abuso potencial de la neurotecnología por parte del gobierno, las agencias públicas o el sector privado”.

Por otra parte, resulta fundamental incidir en la educación y la formación en derechos humanos como herramienta prevención de usos negativos de la tecnología. No solo la formación en derechos sino también la formación tecnológica que, en todo caso, han de abordar cuestiones éticas.

Esta indudable necesidad de formar en derechos y valores se hace patente cuando abordamos los sesgos en la tecnología y su impacto directo y negativo en los derechos a causa de la discriminación algorítmica. En efecto, la lucha contra la discriminación es una constante en la historia de los derechos, por lo que cuando hablamos de discriminaciones en el entorno digital o tecnológico hay que abordar la regulación legal, tecnológica y ética para prevenirla y sancionarla en su caso.

Un breve análisis de la desigualdad y discriminación algorítmica requiere referirse a la privacidad y a los datos personales. Como resulta sobradamente conocido los sistemas de IA ingresan cantidades ingentes de datos, siendo la mayoría datos de carácter personal. Inclusive, algunos de estos dispositivos basados en IA son recolectores masivos de datos personales que se infiltran de manera imperceptible nuestra vida cotidiana, por lo que no todos los datos que se recogen cuentan con conocimiento y consentimiento de las personas interesadas. A mayores, los sistemas de IA procesan de forma sumamente compleja las informaciones personales y obtienen gracias a este tratamiento resultados valiosísimos para el negocio a partir de datos desagregados e incluso anónimos. Por todo ello, la protección de datos es uno de los retos que suelen señalarse en el caso de los sistemas de IA.

Tal como se viene afirmando, la tecnología en general y la IA en particular debe desarrollarse, implementarse y utilizarse desde un enfoque basado en derechos humanos que obliga a colocar a las personas más vulnerables en el centro y tomar decisiones inspiradas en la eficacia de los derechos. Por tanto, todo proceso que implica tratamiento de datos personales debe cumplir la normativa de privacidad y protección de los datos y de los derechos de las personas en general. En concreto las obligaciones establecidas en el Reglamento General de Protección de Datos – Reglamento (UE) 2016/679– (RGPD), así como la Ley de protección de datos española, ya citada. La privacidad supone un espacio de autonomía personal en la sociedad digital.

Son muchas las aplicaciones de IA con impacto en la protección de datos, así las técnicas de reconocimiento fácil y utilización de datos biométricos, el uso de sistemas de IA en los procesos de gestión de datos y de toma de decisiones, el uso de IA con funciones predictivas, la prevención de la discriminación algorítmica o la elaboración de perfiles. Un ejemplo mundialmente conocido fue el caso de Cambridge Analytica, en el que un test de personalidad

llevado a cabo a usuarios de Facebook se usó para inferir perfiles psicológicos de cada uno de ellos. La empresa pudo dirigir comunicaciones personalizadas gracias al conocimiento de las personas a través de técnicas de big data e IA con el objetivo de obtener réditos políticos dirigiendo el voto en un sentido determinado.

El campo de la protección de datos está fuertemente regulado. Podríamos decir que la normativa de la protección de datos está consolidada. Sin embargo, las novedades de la IA han provocado dificultades en la aplicación de determinados principios propios de la protección de datos, que han de estar presentes en los sistemas de IA que emplean datos personales, como por ejemplo la transparencia, el derecho de acceso, el derecho de información. La ética, bajo la fórmula de la prevención del daño y del respeto a los derechos fundamentales debe presidir estas dificultades de aplicación.

Decíamos más arriba que las técnicas de big data y los algoritmos pueden causar lesionar el derecho a no ser discriminado muy relacionado con la protección de las informaciones personales. Concretamente, en términos del Parlamento Europeo, los macrodatos pueden resultar en un tratamiento diferenciado injustificado “y en una discriminación indirecta de grupos de personas con características similares, en particular en lo que se refiere a la justicia e igualdad de oportunidades en relación con el acceso a la educación y al empleo, al contratar o evaluar a las personas o al determinar los nuevos hábitos de consumo de los usuarios de los medios sociales” (Parlamento Europeo, 2017, consideración general decimonovena).

La discriminación se puede producir en distintas fases del proceso de toma de decisiones mediante el uso de datos: desde el momento en que estos son recogidos hasta el de la aplicación del algoritmo por el que se toma la decisión de que se trate.

En un primer momento podría pensarse que estas técnicas presentan una objetividad intachable, pero no es así, o al menos no siempre lo es. Detrás de los datos están las personas que fijan las reglas y criterios de recolección, uso, finalidad, etc, de modo que los datos, en este sentido, también portan subjetividad. En muchas ocasiones el sesgo no será evidente, e incluso puede ocurrir que la persona que trabaja con el sistema en cuestión no esté reflejando en el proceso ideas o valores discriminatorios, pero que el resultado finalmente sí lo sea.

La discriminación algorítmica introduce el concepto de sesgos que penetran voluntaria o inconscientemente en las informaciones con las que se alimentan los algoritmos a través de los cuales deciden o adoptan decisiones las máquinas. Los sesgos son múltiples y variados. Hay sesgos históricos, aprendidos o heredados, otros que se van construyendo con el tiempo a la par que se suceden los cambios sociales. Algunos sesgos son directos, se ven de lejos, otros son sutiles y con consecuencias muy negativas para quienes los sufren: mujeres en el caso de discriminación de sexo, personas mayores en el caso del edadismo, racismo en el caso de personas de otra raza distinta a la blanca, estrato social, religión, y otros muchos.

Los conceptos que se manejan son distintos e incluso se debate sobre si todos los sesgos son negativos y, por lo tanto, inasumibles o si, al contrario, es incluso necesario asumir algunos sesgos para mantener aquellas diferencias que hacen que sea posible la equidad. En definitiva, se necesita que existan porque no se puede tratar igual lo diferente, pues esto no es igualdad.

Los algoritmos y, en consecuencia, los sistemas de inteligencia artificial pueden ofrecer decisiones malas en cuanto que sesgadas, erróneas, o porque perfilan o discriminan, unos efectos o consecuencias que, indefectiblemente, obligan a intervenir y regular. Las

particularidades tecnológicas escapan de la rigidez normativa y, por esto, se recurre a las normas de derecho blando para regular o autorregular. Así, la mayor parte se encuentra en textos, Informes y Cartas (éticas), como las aprobadas por UNESCO o por la institucionalidad latinoamericana o europea empeñadas en regular diversos aspectos “éticos” de la IA. Serían buenos ejemplos de esto, la Declaración de Principios Éticos para la IA de Latinoamérica IA-LATAM para el diseño, desarrollo y uso de la Inteligencia Artificial, que, en su punto octavo, hace referencia expresa a los sesgos: “Evitar los sesgos e impactos injustos en las personas, en particular las relacionadas con características sensibles como la raza, el origen étnico, el género, la nacionalidad, los ingresos, la orientación sexual, la capacidad y las creencias políticas o religiosas”.

En la citada Propuesta de Reglamento de la Comisión Europea del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA (Ley de Inteligencia artificial) de 2021 encontramos en su punto 3.5, sobre el uso de una IA fiable en cuanto a los derechos fundamentales, que se hace mención expresa a la Carta de derechos fundamentales de la UE, y a la promoción de la protección de derechos como el de no discriminación (art.21) y a “minimizar el riesgo de adoptar decisiones asistidas por IA erróneas o sesgadas”.

Como se ha dicho, los sesgos algorítmicos no ocurren de manera espontánea. Los datos no producen sesgos, los algoritmos no producen sesgos, las máquinas no producen sesgos porque los sesgos están en la realidad social y después se reproducen. Hay tres momentos o tres fases en las que se reproducen: cuando se recopilan, cuando se utilizan para entrenar y cuando se toman las decisiones. En la primera porque los datos obtenidos ya llevan con ellos los prejuicios existentes, en la segunda porque cuando se preparan los datos para entrenar al sistema seleccionamos los atributos conforme a nuestros sesgos y, en la última, porque se proyectan los sesgos en las decisiones.

Los algoritmos de IA basados en datos, no producen sesgos pero sí pueden reproducirlos sin la adecuada intervención humana y ello en tres de las fases principales: en la recolección de los datos, porque tales datos recopilados reflejen prejuicios ya existentes; en la preparación de datos de entrenamiento (a la hora de seleccionar y procesar los atributos que le proporcionamos al algoritmo); y en la toma de decisiones (las propuestas y decisiones que se adoptan a lo largo de todo el ciclo de vida del desarrollo inteligente).

El problema ético que se produce en estos casos se entrecruza con un problema técnico que requiere supervisión humana en todas las fases de su ciclo de vida. Más si cabe en los casos que afectan a la Administración pública que debe ser especialmente vigilante con la salvaguarda de los derechos de la ciudadanía, de manera que los programas de IA que se elaboren y apliquen sean comprensibles, transparentes y explicables sin olvidar la necesaria supervisión humana final que haga posible el control y supervisión.

Dada la exigencia de no-discriminación y respeto a la igualdad que se exige a los algoritmos, cabe preguntarse si pueden ser justos, equitativos o imparciales en los sistemas inteligentes y, en su caso, que requisitos o niveles de adecuación resultan exigibles para determinar ese cumplimiento. Nos referimos a exigencias éticas que buscan que los algoritmos no causen daños o que causen los mínimos posibles.

Esto lo anterior nos lleva a la intervención humana, a las responsabilidades exigibles a quiénes intervienen en el ciclo de vida del algoritmo: los datos, medición de resultados para identificar sesgos, definir qué requisitos éticos serán exigibles.

La UNESCO con su “Proyecto de texto de la Recomendación sobre la ética de la Inteligencia Artificial” de 2021 aprobó la primera norma mundial sobre la ética de la inteligencia artificial. En su ámbito de actuación número seis, sobre “género”, enumera un conjunto de deberes que competen a los Estados Miembros para lograr la igualdad de género en los sistemas de IA. Específicamente y con relación a los sesgos, en el punto 90 se dice: [...] velar por que los estereotipos de género y los sesgos discriminatorios no se trasladen a los sistemas de IA, sino que se detecten y corrijan de manera proactiva. [...].

El texto tiene la virtualidad de que no se limita a tratar el género como un problema de sesgos al que hay que dar una solución, sino que va más allá, estableciendo como deber de los Estados que las normas relativas a la IA y a la automatización, de una manera transversal, tengan un enfoque de género. El reto no es exclusivamente el de detectar y eliminar los sesgos algorítmicos sexistas sino aprovechar el potencial que brinda la IA para promover un enfoque de género, impulsando la efectiva igualdad de género-.

España ha remitido en 2020, su Estrategia para la IA, en la que, siguiendo las orientaciones de Informes y Documentos de la Unión Europea, ya incluye referencias explícitas e implícitas al género -el desafío social 1 es “Reducir la brecha de género del ámbito de la IA en empleo y liderazgo”: “[...] la igualdad de género ha de ser uno de los objetivos transversales de la presente estrategia”- lo que ya constituye un adecuado punto de partida para transitar hacia una donde el género no sea un factor de discriminación. La Carta de derechos digitales, adoptada en julio de 2021 por el Gobierno de España, y que se inscribe en el contexto de la Estrategia Española Nacional de Inteligencia Artificial -ENIA- de 2020, aunque carece de efectos normativos, en su punto seis, apuesta por “Establecer un marco ético y normativo que refuerce la protección de los derechos individuales y colectivos, a efectos de garantizar la inclusión y el bienestar social”.

Aunque la Carta no tiene carácter normativo, se presenta como la línea a seguir en la legislación y políticas públicas. La Carta hace especial incidencia con respecto a los derechos en entornos específicos, como es el caso los derechos ante la Inteligencia artificial. En el Capítulo XXV, titulado “Derechos ante la inteligencia artificial”, en el apartado primero, defiende un enfoque centrado en la persona y en su inalienable dignidad; en el apartado segundo, con respecto al desarrollo y ciclo de vida de los sistemas de inteligencia artificial, en su apartado a) preceptúa el derecho a la no discriminación:

“Se deberá garantizar el derecho a la no discriminación cualquiera que fuera su origen, causa o naturaleza, en relación con las decisiones, uso de datos y procesos basados en inteligencia Artificial”. Si la finalidad planteada es clara, el problema radica en el cómo conseguir esa no discriminación. De ahí que, continúa en el apartado b: “Se establecerán condiciones de transparencia, auditabilidad, explicabilidad, trazabilidad, supervisión humana y gobernanza. En todo caso, la información facilitada deberá ser accesible y comprensible”. Y se completa con lo establecido en el apartado tercero, según el cual “Las personas tienen derecho a solicitar una supervisión e intervención humana y a impugnar las decisiones automatizadas tomadas por sistemas de inteligencia artificial que produzcan efectos en su esfera personal y patrimonial”.

Se deben adoptar tanto medidas de índole política y legislativa, como medidas propiamente tecnológicas, con una IA que garantice el escrupuloso respeto de los derechos fundamentales y que resulte “fiable” -fiabilidad que exigen documentos y Directrices éticas de diversos organismos e instituciones, como la UNESCO y la Unión Europea-. Los principios rectores sobre la IA de alto riesgo -el ámbito jurídico lo es en cuanto afecta a derechos fundamentales-

requiere transparencia, rendición de cuentas, supervisión humana y ausencia de sesgo y de discriminación. Para ello, los algoritmos deben ser auditables, transparentes y explicables, tal y como se exige en varios documentos internacionales e instrumentos normativos, lo que a su vez requiere establecer un sistema de control y de auditoría. No basta con establecer cómo deben ser los algoritmos sino verificar que, efectivamente, se están cumpliendo los requisitos establecidos.

Así pues, la tecnología solo tiene sentido si el objetivo o finalidad es servir a las personas y mejorar a través de procesos eficientes la calidad de los servicios y la toma de decisiones. Los datos de calidad deben ser el nutriente del algoritmo que alimenta a la máquina que arrojará decisiones o ayudas para tomarlas y la calidad está directamente relacionada con el buen o mal hacer de las personas.

Una formación ética, es decir, en valores y principios dimanantes de los derechos humanos fundamentales es esencial para hacer realidad el objetivo de progresar de forma ética y eficiente que permita la fiabilidad y seguridad tecnológica que demandan las personas que serán las que usen la tecnología.