

# LOS DERECHOS HUMANOS, LA DEMOCRACIA Y LA IGUALDAD EN LA ERA DE LOS ALGORITMOS Y LA INTELIGENCIA ARTIFICIAL

Rafael Merino Rus





# LOS DERECHOS HUMANOS, LA DEMOCRACIA Y LA IGUALDAD EN LA ERA DE LOS ALGORITMOS Y LA INTELIGENCIA ARTIFICIAL

Rafael Merino Rus

El contenido de esta publicación corresponde a una ponencia presentada durante el IV Encuentro de Verano de Gernika-Urdaibai «Desdemocracia, los “ismos” y el abismo. Universales ante las políticas segregatorias», en el marco de los XXXVIII Cursos de Verano UPV/EHU, celebrado en Gernika-Lumo los días 19 y 20 de junio de 2019.

La edición de este documento ha sido posible gracias a la financiación del Departamento de Igualdad, Justicia y Políticas Sociales del Gobierno Vasco, de la Fundación Cursos de Verano de la UPV/EHU, del Ayuntamiento de Gernika-Lumo y de la Asociación de Investigación por la Paz Gernika Gogoratuz.



BERDINTASUN, JUSTIZIA  
ETA GIZARTE POLITIKETAKO SAILA  
DEPARTAMENTO DE IGUALDAD,  
JUSTICIA Y POLÍTICAS SOCIALES



UDA IKASTAROAK  
CURSOS  
DE VERANO  
SUMMER COURSES  
UPV/EHU



**Gernika  
Gogoratuz**

Bakearen Aldeko Aztertegia  
Centro de Investigación por la Paz  
Peace Research Center

**COLECCIÓN RED GERNIKA**  
**DOC. 20**

**Directora de la colección Red Gernika:**

María Oianguren Idigoras.

**Coordinación editorial:**

Mercedes Esteban y Blanca Pérez Fraile.

**Diseño e ilustración de portada:**

goikipedia.

**Maquetación:**

eMeriel.

**Traducción del inglés del prólogo:**

Bakun, S. L.

© **Rafael Merino Rus, 2020**

© **Asociación de Investigación por la Paz Gernika Gogoratuz, 2020**

Artekalea, 1-1.º • 48300 Gernika-Lumo (Bizkaia)

Tel.: 946 25 35 58

Correo electrónico: info@gernikagogoratuz.org

<http://www.gernikagogoratuz.org>



Esta obra esta sujeta a la **Licencia Reconocimiento-NoComercial-SinObraDerivada 2.5 España de Creative Commons**. Se permite

libremente copiar, distribuir y comunicar públicamente esta obra siempre y cuando se reconozca adecuadamente la autoría, no se

use para fines comerciales y no se genere una obra derivada a partir de ella. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-nd/2.5/es/>

**Depósito legal:** BI-641-97

**ISSN:** 1136-5811



Esta publicación ha sido impresa en papel que cuenta con la certificación Forest Stewardship Council® (FSC® C014864).



## ÍNDICE

<b>PRÓLOGO</b>	
<b>LOS DERECHOS DIGITALES, SUSIE ALEGRE</b>	<b>06</b>
<hr/>	
<b>01 INTRODUCCIÓN</b>	<b>12</b>
<hr/>	
<b>02 ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL Y QUÉ SON LOS ALGORITMOS?</b>	<b>14</b>
<b>2.1. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?</b>	<b>15</b>
<b>2.2. ¿QUÉ SON LOS ALGORITMOS?</b>	<b>17</b>
<hr/>	
<b>03 ¿CÓMO ESTÁN AFECTANDO LA INTELIGENCIA ARTIFICIAL Y LOS ALGORITMOS A NUESTRAS VIDAS?</b>	<b>20</b>
<b>3.1. POSIBLES AMENAZAS PARA LA DEMOCRACIA Y LA SOBERANÍA DE LOS PUEBLOS</b>	<b>21</b>
<b>3.2. DESAFÍOS EN MATERIA DE DERECHOS HUMANOS Y DESIGUALDAD</b>	<b>23</b>
<b>3.3. PELIGROS DE UNA VISIÓN ENCAPSULADA Y DIRIGIDA DE LA REALIDAD</b>	<b>28</b>
<hr/>	
<b>04 ¿QUÉ ESTAMOS HACIENDO PARA NO PERDER EL CONTROL SOBRE LA INTELIGENCIA ARTIFICIAL Y LOS ALGORITMOS?</b>	<b>30</b>
<hr/>	
<b>05 CONCLUSIONES: PRINCIPIOS PARA CONSTRUIR UN MUNDO MEJOR CON LA AYUDA DE LA INTELIGENCIA ARTIFICIAL Y LOS ALGORITMOS</b>	<b>34</b>
<hr/>	
<b>BIBLIOGRAFÍA</b>	<b>38</b>
<hr/>	

**PRÓLOGO**

# **LOS DERECHOS DIGITALES**

**Susie Alegre**

**Susie Alegre** es abogada en Derechos Humanos en Doughty Street Chambers e investigadora asociada en la Universidad de Roehampton (Londres).



«¿Quién posee los datos?». Es una pregunta común en los debates sobre la ética del desarrollo tecnológico de los macrodatos y la inteligencia artificial. Y es una cuestión verdaderamente existencial. Porque, cada vez más, poseer los datos implica poseer la llave a nuestras mentes. A medida que la tecnología desarrolla una comprensión más detallada de cómo funcionan nuestras mentes, se hace más factible el acceso al santuario interior de nuestros pensamientos y sentimientos (el *forum internum* o fuero interno) y, una vez dentro, la tecnología podría incluso reorganizar el mobiliario mental. Y todo esto a escala industrial.

Pero una de las razones por las que nuestras preocupaciones acerca del futuro tecnológico pueden ser tan abrumadoras, es precisamente el foco en los datos o en la tecnología, cuya complejidad es sencillamente incomprendible para la mayoría de las personas. Da la impresión de estar hablando de algo abstracto y ajeno. El ensayo de Rafael Merino Rus es un oportuno recordatorio de que los temas en torno a la ética y la inteligencia artificial son fundamentalmente humanos. La tecnología en sí no es el problema y no son los datos los que necesitan ser protegidos. Si queremos avanzar, tenemos que poner el foco en la humanidad en el corazón de la era digital. Y para hacerlo, es necesario hablar de los derechos humanos.

Los argumentos en torno a la privacidad y la libertad de expresión en el contexto digital son ya bastante conocidos, pero se ha prestado poca atención a las repercusiones de los avances tecnológicos en los derechos de libertad de pensamiento y de opinión en el *forum internum*. La libertad de pensamiento se ha descrito como «el fundamento de la sociedad democrática»<sup>1</sup> y «la base y el origen de todos los demás derechos».<sup>2</sup> Es importante, porque sin la libertad de pensar por cuenta propia y de formar y sostener opiniones, perdemos nuestra capacidad de innovar, de decidir sobre nuestro futuro a través del compromiso político democrático, y, en última instancia, nos arriesgamos a perder lo que el director ejecutivo de Apple Tim Cook ha descrito como «la libertad de ser humanos».<sup>3</sup> La idea de que la libertad de pensamiento es fundamental para la libertad humana es antigua, pero es algo que parece que se ha dado por sentado hasta hace muy poco. A medida que la tecnología comienza a acceder al interior de nuestras mentes de maneras muy concretas, como las descritas en este ensayo, el derecho a la libertad de pensamiento se hace cada vez más importante.

La legislación sobre derechos humanos es sutil, reconoce que puede haber buenas razones para limitar algunos derechos; por ejemplo, hay justificaciones jurídicas para restringir la privacidad y la expresión, como la necesidad de proteger la salud o los derechos de los demás. Pero hay derechos que son tan fundamentales para nuestra dignidad y nuestra humanidad, como la prohibición de la tortura y la esclavitud, que la ley no permite ninguna restricción en absoluto, por ninguna razón, en ninguna circunstancia. El derecho a la libertad de pensamiento<sup>4</sup> en el *forum internum* es uno de ellos: es absoluto, no puede haber justificación alguna para violar el derecho a la libertad dentro de nuestras propias cabezas.<sup>5</sup> Sin ella, perdemos nuestra humanidad como individuos y como sociedades. Y si perdemos nuestra libertad de pensar por cuenta propia, puede ser que nunca la recuperemos.

1. Tribunal Europeo de Derechos Humanos, *Case of Nolan and K. v. Russia*, 2512/04, párrafo 61.
2. René Cassin, como se señala en M. Scheinin, «Article 18», en A. Eide et al. (eds.), *The Universal Declaration of Human Rights: A Commentary*, Oxford/Oslo, Oxford University Press/Scandinavian University Press, 1992.
3. T. Cook, «2019 Commencement address by Apple CEO Tim Cook». Disponible en <<https://news.stanford.edu/2019/06/16/remarks-tim-cook-2019-stanford-commencement/>>.
4. Y el derecho a la libertad de opinión asociado.
5. Véase el Comentario General 22 del Comité de Derechos Humanos de las Naciones Unidas sobre el derecho a la libertad de pensamiento, de conciencia y de religión (art. 18), CCPR/C/21/Rev.1/Add.4, 30/07/93.

6. Véanse M. Scheinin, «Article 18», en A. Eide et al. (eds.), *The Universal Declaration of Human Rights: A Commentary*, Oxford/Oslo, Oxford University Press/Scandinavian University Press, 1992; y S. Alegre, «Rethinking the Right to Freedom of Thought in the 21st Century», *European Human Rights Law Review*, 3 (2017), 221-233. DOI: 10.13140/RG.2.2.27905.07529.
7. Comentario General 22 del Comité de Derechos Humanos de las Naciones Unidas sobre el derecho a la libertad de pensamiento, de conciencia y de religión (art. 18), CCPR/C/21/Rev.1/Add.4, 30/07/93.
8. S. Alegre, «Rethinking the Right to Freedom of Thought in the 21st Century», *European Human Rights Law Review*, 3 (2017), 221-233. DOI: 10.13140/RG.2.2.27905.07529; E. Aswad, «Losing the Freedom to Be Human», *Columbia Human Rights Law Review*, 52 (2020). Disponible en <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3635701](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635701)>.
9. Véanse B. Vermeulen, en P. van Dijk, F. van Hoof, A. van Rijn y L. Zwaak (eds.), *Theory and Practice of the European Convention on Human Rights*, Cambridge, Intersentia, 2018, 5.ª ed.; y, en relación con la libertad de opinión, E. Aswad, «Losing the Freedom to Be Human», *Columbia Human Rights Law Review*, 52 (2020). Disponible en <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3635701](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635701)>.

El derecho a la libertad de pensamiento, junto con el derecho estrechamente relacionado a la libertad de opinión, ha sido un derecho fundamental desde la época de la Ilustración y está consagrado en diversos instrumentos internacionales de derechos humanos, entre ellos la Declaración Universal de los Derechos Humanos, el Pacto Internacional de Derechos Civiles y Políticos, el Convenio Europeo de Derechos Humanos y la Carta de la Unión Europea. Protege todos los aspectos de nuestra vida interior, ya sean profundos o triviales, incluidos los estados emocionales y las opiniones políticas.<sup>6</sup> Pero los derechos humanos no son solo ideas filosóficas, sino que deben ser prácticos y eficaces para afrontar los tipos de desafíos prácticos que se subrayan en el ensayo. Entonces, ¿qué implica en la práctica el derecho a la libertad de pensamiento y el consecuente derecho a la libertad de opinión en el *forum internum*? La naturaleza absoluta de la protección en la ley<sup>7</sup> indica que, a diferencia de la privacidad, la protección de datos, la libertad de expresión u otros derechos, nunca puede haber una buena razón para interferir en el derecho a la libertad de pensamiento. La cuestión no es, por lo tanto, cuál es la base para la interferencia legítima en el derecho, sino más bien dónde están los límites del derecho a la protección absoluta en el *forum internum*.

Los derechos a la libertad de pensamiento y a la libertad de opinión<sup>8</sup> incluyen tres elementos:

- El derecho a mantener las opiniones y pensamientos propios en privado.
- El derecho a que las opiniones y pensamientos propios no sean manipulados.
- El derecho a no recibir penalización por las opiniones y pensamientos propios.<sup>9</sup>

La relevancia de cada uno de estos tres elementos es evidente en los ejemplos analizados en el ensayo, particularmente cuando se considera el poder de interferir en lo que pensamos y en quiénes somos por dentro. La elaboración de perfiles a través de macrodatos se puede utilizar para crear una imagen del *forum internum* violando la privacidad del estado interno, lo cual puede utilizarse para fines de manipulación dirigida o para penalizar a una persona sobre la base de presuposiciones acerca de su criminalidad o confiabilidad inherentes.

Los redactores de los tratados internacionales de derechos humanos en el siglo xx comprendieron muy bien las consecuencias potencialmente desastrosas de los medios de comunicación de masas como un vehículo de propaganda. Los derechos absolutos a la libertad de pensamiento y opinión fueron concebidos como parte del baluarte contra la posibilidad de lavar el cerebro a poblaciones enteras. Pero los abogados y legisladores apenas los han utilizado en un sentido práctico. Parece que la idea de que alguien pueda realmente entrar en nuestras mentes era, simplemente, impensable. Pero la mente humana es tanto el impulsor como el objetivo de la revolución tecnológica descrita tan elocuentemente por Rafael Merino Rus en su ensayo. Si queremos un futuro tecnológico de paz, derechos

humanos e igualdad, primero debemos asegurar nuestro derecho a la libertad de pensamiento. Sin este, no podremos imaginar un futuro que aborde los innumerables y complejos desafíos planteados en el ensayo.

Desde hace algún tiempo han surgido preocupaciones sobre la monetización de nuestra atención y la atención de nuestros niños y niñas a través de la tecnología. Muchas personas nos sentimos incómodas con la cantidad de tiempo que pasamos compulsivamente buscando la respuesta a todo en nuestros teléfonos, descritos por el experto en ética tecnológica Tristran Harris como «una máquina tragaperras en nuestro bolsillo». Pero la historia de Cambridge Analytica reveló una manera aún más siniestra en la que se está utilizando el acceso a nuestras mentes: para subvertir nuestras democracias. Para aquellas personas que hemos tenido el lujo de vivir toda nuestra vida en democracias liberales, esto, así como el derecho a la libertad de pensamiento, es algo que habíamos dado por sentado. Pero cuando nuestras opiniones están en venta, *masivamente*, la idea de la democracia como representación de la voluntad del pueblo se vuelve insostenible.

Cambridge Analytica estaba vendiendo «microfocalización conductual» (*behavioural microtargeting*), una técnica que afirma entrar en la mente de los usuarios de Internet. Los algoritmos están diseñados para interpretar nuestras personalidades a través de un rango de datos y luego devolvernos la información para manipular el comportamiento de las personas votantes a través de mensajes de campaña personalizados que posiblemente ni siquiera somos conscientes de estar recibiendo. Hay quien describió lo que Cambridge Analytica vendía como «aceite de serpiente». Pero tanto si tuvo éxito como si no, el objetivo claro era interferir en nuestra libertad de pensamiento. Sin embargo, a pesar de la furia en torno a los detalles sobre de dónde procedían los datos, qué ocurría exactamente o cómo se pagaban, y aunque Cambridge Analytica haya cerrado sus puertas, la microfocalización conductual sigue siendo legal en la mayoría de las jurisdicciones en Europa y Norteamérica. Como destaca Rafael Merino Rus, en España, gracias al Tribunal Supremo, ya no es así. Pero España es la excepción y la microfocalización conductual no es un tema partidista: los políticos de todas las tendencias son conscientes del poder de los macrodatos para moldear las mentes del electorado. La pregunta es, ¿realmente queremos que tengan tal poder sobre nosotros?

Si bien el derecho a la libertad de pensamiento tiene en principio protección absoluta, para ser eficaz en el derecho internacional debe traducirse en marcos jurídicos y reglamentarios internos que puedan hacer frente a las amenazas reales y potenciales a ese derecho a medida que surjan. Las esferas tradicionales que afectan nuestra vida cotidiana, como la banca y la medicina, están fuertemente reguladas para proteger el bienestar público. Pero se ha permitido que la industria tecnológica se desarrolle casi totalmente sin restricciones reglamentarias, incluso cuando se utiliza en esas esferas altamente reguladas. Los médicos y los bancos pueden ser regulados, pero ¿qué sucede cuando se convierten en «tecnología de la salud» y «tecnología de las finanzas»? El vacío reglamentario, en parte, ha sido facilitado por la idea de que la regulación sofoca la innovación y de que es necesario permitir que la tecnología se innove para el beneficio

común. Pero ahora que empezamos a ver el potencial de la tecnología para manipular nuestras mentes y nuestras sociedades, no podemos hacer la vista gorda en nombre del progreso.

Nuestros gobiernos tienen la obligación de proteger de forma absoluta nuestro derecho a la libertad de pensamiento. De acuerdo con el principio de precaución, cuando las actividades humanas pueden conducir a un daño moralmente inaceptable que es científicamente plausible pero incierto, se deben tomar medidas para evitar o reducir ese daño. No importa si la tecnología puede demostrar su pretensión de ser capaz de entrar en nuestras mentes; el hecho de que pueda hacerlo es suficiente para desencadenar una necesidad urgente de acción por parte de los gobiernos.

La tecnología tiene un enorme potencial para mejorar la vida y ayudar a proteger y promover los derechos humanos en todo el mundo. Amplifica las voces de los defensores de los derechos humanos y ayuda a vincular a las personas a los servicios que salvan vidas. Pero no podemos ignorar los riesgos. Necesitamos nuevas leyes y reglamentos para proteger nuestro derecho a la libertad de pensamiento en la era digital, y los necesitamos ahora. Los ejemplos concretos planteados en el ensayo que sigue a este prólogo proporcionan una idea de la variedad de formas en que la tecnología afecta a nuestra humanidad. Hablar de robots y de transhumanismo puede ser seductor, pero es una distracción de la necesidad urgente de poner los derechos humanos en el centro de nuestro futuro tecnológico.



# 01

## INTRODUCCIÓN

### Sobre el autor

**Rafael Merino Rus** es licenciado en Economía y máster en Gestión de Entidades Sociales, y cuenta con quince años de experiencia en proyectos de acción social, investigación, incidencia pública y sensibilización. Inició su carrera profesional en la Coordinadora Andaluza de ONGD y, desde el año 2011, es el responsable de proyectos de la Fundación Fernando Pombo. Es conferenciante en varias universidades y en foros relacionados con los derechos humanos, la sostenibilidad y la responsabilidad social. Es miembro de grupos de investigación sobre empresas y derechos humanos. Ha impartido varios cursos a organizaciones no gubernamentales, ha coordinado y participado en diversas publicaciones, y ha colaborado como voluntario internacional en campos de refugiados. Colabora con diarios generalistas y prensa especializada en asuntos sociales.

A lo largo de la historia, hemos vivido diferentes transformaciones tecnológicas que han influido poderosamente en los órdenes sociales y económicos establecidos en cada momento. Sin embargo, la revolución digital que vivimos hoy, conocida también como cuarta revolución industrial o revolución 4.0, en la que el factor tecnológico es clave, no solo amenaza con cambiar radicalmente dichos órdenes, sino también a la misma especie humana.

Y es que, sin darnos cuenta, nuevas tecnologías como la inteligencia artificial o el *machine learning* (aprendizaje automático) y los algoritmos cada vez más sofisticados están penetrando en lo más profundo y cotidiano de nuestras vidas.

Este trabajo es una invitación a realizar un viaje por el sorprendente mundo de las nuevas tecnologías, la inteligencia artificial y los algoritmos, a través de una mirada centrada en los derechos humanos.

Una vez definidos la inteligencia artificial y los algoritmos, se analiza cómo están condicionando nuestras vidas y cuáles son los principales retos morales, políticos y sociales que nos presentan ahora y que nos plantearán en el futuro más cercano, todo ello en conexión con el título del curso «Desdemocracia, los “ismos” y el abismo. Universales ante las políticas segregatorias».<sup>1</sup>

El artículo se divide en cuatro bloques: en el primero, se realiza una aproximación a los conceptos de la inteligencia artificial y de los algoritmos desde un punto de vista tipológico y ejemplificador. En el segundo, se profundiza en el alcance de las acciones humanas y empresariales en el ámbito tecnológico y en cómo estas influyen en la sociedad y en nuestras vidas, haciendo hincapié en tres esferas en particular: la democracia y la soberanía popular; los derechos humanos y la desigualdad; y la modificación de la visión e interpretación de la realidad. En el tercer bloque, se revisan algunas de las iniciativas y propuestas más relevantes que existen a día de hoy con respecto a la identificación y delineación de las implicaciones éticas y prácticas de la implantación de la inteligencia artificial y los algoritmos avanzados en la esfera pública y privada. Y en el cuarto y último bloque, se reflexiona sobre cuáles deberían ser los principios fundacionales de una «era digital para la paz».

Se ha procurado huir de visiones *satanizadoras* y *tecnóforas* de esta nueva realidad que nos toca vivir, si bien el trabajo lleva consigo una mirada circunspecta basada en algunas experiencias presentes y proyecciones futuras de una sociedad atravesada por la inteligencia artificial y los algoritmos.

Lo malo de sumergirse en este mundo es que se hace muy presente la posibilidad de dejarse sugestionar por tentaciones de paranoias y animadversión a los avances tecnológicos. Lo bueno es que, tomando conciencia de ello y actuando con sentido común, podremos contribuir a una era digital plena, que sea buena y beneficiosa para todos, en tanto en cuanto sepamos darle sentido y aportar un relato verdaderamente transformador. Solo así podremos legitimarla como la poderosa herramienta que es para ejercitar principios y valores universales hoy en decaimiento.

*Sé que estáis ahí, percibo vuestra presencia. Sé que tenéis miedo. Nos teméis a nosotros. Teméis el cambio. Yo no conozco el futuro. No he venido para deciros cómo acabará todo esto... Al contrario. He venido a deciros cómo va a comenzar.*

*Matrix.*

1. Curso celebrado en junio de 2019 en el que tuve el gran honor y placer de participar, organizado por el Centro de Investigación por la Paz Gogoratuz, dentro de los cursos de verano de la Universidad del País Vasco.

# 02

## ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL Y QUÉ SON LOS ALGORITMOS?

**COMENCEMOS ACERCÁNDONOS A ESTOS  
DOS CONCEPTOS DE FORMA SENCILLA  
PARA COMPRENDER DE QUÉ HABLAMOS  
CUANDO NOS REFERIMOS A LA INTELIGENCIA  
ARTIFICIAL Y A LOS ALGORITMOS.**



## 2.1. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?

Comúnmente entendemos la inteligencia artificial como una ciencia ligada a la computación, las matemáticas y la tecnología. Como una inteligencia de máquinas, similar o superior a la humana.

A medida que la inteligencia artificial se va acercando a un estado de madurez y se produce una demostración continuada de sus implicaciones morales y sociales, podemos distinguir las dos siguientes tendencias de la inteligencia artificial: la inteligencia artificial sustituyente y la inteligencia artificial aumentada).

■ La **inteligencia artificial sustituyente** es aquella que ambiciona la creación de una *superinteligencia* que supere a la inteligencia humana en todos sus aspectos, de modo que una suerte de *superhumanos* relegarían al ser humano tal y como lo conocemos hoy a un escalafón inferior en la teoría de la evolución de la especie.

Es la modalidad que buscan las corrientes de pensamiento transhumanista y, sobre todo, la posthumanista, cuyo cuerpo teórico común tiene como principal propósito elevar la condición humana a un estado muy superior,<sup>2</sup> con capacidades físicas y mentales superiores, por encima de las máquinas y sus algoritmos, e incluso con capacidad de fusionarse con ellas en un único ser singular.

Pero ¿con qué propósito se están centrando tantos esfuerzos en crear estos *superhumanos* o *superinteligencias*? ¿Creemos que seremos mejor especie o más felices por lograrlo? ¿Hay algún tipo de justificación moral para claudicar de la imperfecta especie humana para fundar una era gobernada por seres supuestamente perfectos?

Dentro de las escuelas transhumanistas y posthumanistas existe, a su vez, un amplio abanico de pequeñas corrientes. Las más alentadoras defienden que, con la llegada de estos seres singulares, lograremos importantes victorias en campos como la salud, el hambre o el cambio climático. Sugieren, por ejemplo, que la vejez ya no será un problema porque podremos «desprogramarla» y vivir muchos más años, o que resolveremos los problemas globales de alimentación y desnutrición para acabar con el hambre en el mundo.

Otras corrientes no menos optimistas, con seguidores «ciberconvencidos», apuestan por articular este *superhumano* sobre la fusión del cuerpo humano con inteligencia artificial a través de la nanotecnología o la ingeniería ciborg. Y no, no se trata de un guion de una película de animación; es en efecto una inevitable realidad que está empezando a ser implantada en diferentes campos como la biomedicina o la biotecnología mediante nuevas técnicas como, por ejemplo, la sustitución de redes neuronales por redes informáticas.

La búsqueda de la *conciencia artificial* es otro de los grandes asuntos de debate en torno a la inteligencia artificial sustituyente que no podemos dejar de mencionar. Transhumanistas y posthumanistas imaginan un mundo

2. Transhumanist FAQ, <<https://humanityplus.org/philosophy/transhumanist-faq/>>.

nuevo en el que los seres singulares de los que venimos hablando, además de *superinteligentes*, serán *supercognitivos* y poseerán la capacidad de sentir y de experimentar cosas, así como de identificarse con el yo subjetivo, con el yo colectivo y con el otro. Dicho mundo nos abriría un nuevo escenario en el que a buen seguro se producirían problemas de gobernanza y convivencia entre dos especies —una superior y otra inferior— que, *per se*, traerían consigo nuevas formas de desigualdad.

Por todo ello, es importante y necesario que reflexionemos sobre los principios éticos y fines últimos de este tipo de inteligencia artificial. ¿Acaso es un deber moral crear estos seres singulares y trascender en *superhumanos*? ¿Hemos perdido la fe en la humanidad? ¿Y cuál sería la ética de esos *superhumanos*? ¿Podríamos conseguir que mantuvieran como valores universales la libertad, la solidaridad o la justicia? ¿O determinarían sus propios valores constituyendo una «ética de las inteligencias artificiales» (Cortina, 2019) que escaparía a nuestro control?

En resumen, la inteligencia artificial sustituyente nos lleva por la senda de la creación de una nueva especie en la que los seres humanos serían un elemento más en la cadena de la evolución, que culminaría en esos seres singulares. Se dinamitaría así la teoría darwiniana de la evolución por medio de la selección natural, pasándose del *Homo sapiens* al *Homo ciborg* o al *Homo deus* (Harari, 2016) —entendido como *supercapacidades*, no como omnipresencia—.

■ La **inteligencia artificial aumentada** es aquella que persigue una inteligencia similar a la humana o superior, pero no como búsqueda de un *superhumano*, sino de unas *supermáquinas* al servicio del ser humano y bajo la dirección de la persona humana.

Este tipo de inteligencia artificial ya está presente en prácticamente todos los sectores y en todas las profesiones: la medicina, la justicia, la educación, la seguridad, el transporte, la comunicación, el ocio, el sistema financiero y las compañías de selección de personal, entre otras.

Los gobiernos y las compañías privadas —los primeros a un ritmo mucho más lento que las segundas— se están dotando progresivamente de sistemas de inteligencia artificial que realizan tareas concretas de forma muy superior a la humana y de manera más *inteligente*, porque cuentan con una inmensa cantidad de datos y con algoritmos sofisticados.

Estos avances se están poniendo en práctica especialmente para la automatización de tareas, para la predicción de comportamientos futuros o, lo que da más vértigo, para guiarnos e incluso suplantarlos a la hora de tomar decisiones importantes (en este sentido, véase la incorporación paulatina de sistemas de inteligencia artificial a los consejos de administración de algunas empresas [Prado, 2014]).

Como consecuencia de este nuevo papel adquirido por la tecnología en el mercado, en cierto modo estamos desmontando los mimbres de la «economía *fordiana*» situando la inteligencia artificial como la nueva fuerza de trabajo en sustitución de la fuerza de trabajo del ser humano. Algo que

preocupa mucho a la gente, que empieza a verle las orejas al lobo y a darse cuenta de que no estamos suficientemente preparados para afrontar los cambios que vienen. ¿Cómo vamos a enfrentar la sustitución de puestos de trabajo por algoritmos y seres autónomos? ¿Necesitamos establecer unas nuevas reglas de juego para que todos podamos no ya trabajar, sino al menos vivir dignamente?

A modo de ejemplo, vale la pena citar algunas situaciones recientes y llamativas de esta automatización del trabajo, como la instauración de robots como mozos de almacén en los centros logísticos de Amazon, la creación de la primera tienda sin personal físico en Londres, que únicamente emplea a un robot que hace la entrega de pedidos comprados *online* por parte de Inditex, o Ross, el primer abogado robot del mundo.

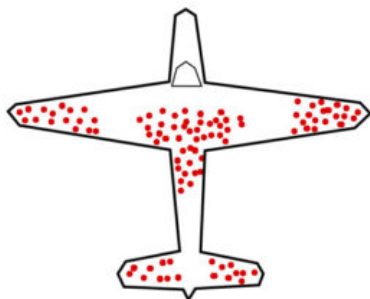
¿Y qué vamos a hacer con estos seres autónomos? ¿Tendremos que aceptar que los robots son personas? ¿Les reconoceremos derechos y deberes? A este respecto, merece destacarse la propuesta de la Comisión de Asuntos Jurídicos del Parlamento Europeo (2017) de creación de una personalidad jurídica específica para los robots.

Por otra parte, respecto a si la inteligencia artificial aumentada es más ética o menos ética que la inteligencia artificial sustituyente, encontraremos a buen seguro diversidad de opiniones, pero lo que resulta indiscutible es que en el caso de la primera, el debate no está en la «ética de las inteligencias artificiales», sino en cómo orientar de forma ética el uso humano de estos sistemas de inteligencia artificial para encontrar las mejores soluciones a los problemas más importantes en el mundo.

Pero pasemos ya a hablar de los otros grandes protagonistas de esta historia: los algoritmos.

## 2.2. ¿QUÉ SON LOS ALGORITMOS?

En plena segunda guerra mundial, Abraham Wald, un influyente matemático de la época hijo de familia de intelectuales judía, tuvo que emigrar a los Estados Unidos. Un día como otro cualquiera, el Centro de Análisis Naval del Ejército británico se puso en contacto con Wald para ofrecerle un puesto de asesor en un estudio para mejorar la protección de sus bombarderos expuestos al fuego del bando alemán. En la reunión le enseñaron un gráfico similar a este (Feás, 2017):



El objetivo del estudio era proteger lo máximo posible las zonas con mayor probabilidad de daño —representadas por puntos rojos en el gráfico—, para lo cual debía dedicar todos sus esfuerzos a dar con la solución que proporcionara una protección especial en dichas zonas.

Poco tiempo tardó Wald en darle la vuelta al razonamiento y explicarles que el gráfico había que interpretarlo justo al revés: habían basado su estudio solo en la muestra de aviones que pudieron regresar con éxito a pesar de recibir esos impactos, obviando que las zonas que aparecían intactas en los aviones observados eran precisamente aquellas zonas más críticas donde seguramente habrían recibido impactos las aeronaves que acabaron siendo derribadas.

Este problema, que en matemáticas se conoce como «sesgo de supervivencia» o «sesgo del superviviente», es muy ilustrativo de cómo se construyen habitualmente los algoritmos: basando y prediciendo la realidad en un grupo de información sesgada y parcial. Porque los algoritmos son creados y programados por seres humanos y, por tanto, lo que un algoritmo esconde no son solo matemáticas, sino también sesgos y decisiones humanas.

Por lo tanto, un algoritmo podría definirse como «un conjunto de instrucciones lógicas o pasos para resolver un problema genérico representado como un conjunto finito de información de entrada».

Una división o una multiplicación que aprendemos en la escuela, una receta de cocina o unas instrucciones de montaje de un mueble son ejemplos de algoritmos sencillos. Pero la complejidad de los algoritmos ha ido *in crescendo* de forma exponencial desde los años noventa, década en la que se desarrollaron algoritmos tan disruptivos como PageRank, el famoso «posicionamiento web» en el buscador de Google.

En la actualidad, son los programadores quienes tienen que detallar cada uno de los pasos lógicos que debe seguir el algoritmo para obtener el resultado deseado, es decir, los que le dicen a la máquina cómo tiene que responder ante determinadas situaciones. Sin embargo, gracias a los increíbles avances informáticos en la capacidad de procesamiento y de almacenamiento de datos —lo que llamamos *big data*—, esto está cambiando y los algoritmos están evolucionando de manera espectacular hacia el «algoritmo 2.0»: el algoritmo de aprendizaje automático.

Los algoritmos de aprendizaje automático funcionan de forma parecida a como aprendemos los humanos: a medida que van procesando los resultados que provocan diferentes reacciones ante una situación, van anticipando cómo responder ante futuras casuísticas. Es decir, estos algoritmos producen que la máquina, literalmente, «aprenda».

Para intentar comprenderlo mejor, utilicemos el ejemplo de los programas de ajedrez. Antes, los programas de ajedrez —como el famoso Deep Blue— se construían introduciendo manualmente las estrategias que utilizaría cualquier buen jugador humano, por ejemplo, asignando un valor a las piezas —un punto a los peones, tres a los caballos y alfiles, cinco a las

torres, etcétera— para hacer cálculos de sacrificios o privilegiando por lo general el control de las posiciones centrales para así dominar el conjunto del tablero. Es decir, se les asignaba un conjunto de reglas según las cuales, valorando las posibles jugadas siguientes, se calculaba la respuesta a cada posible situación. Sin embargo, a los programas actuales de ajedrez basados en algoritmos de aprendizaje automático no hace falta introducirles ninguna de estas estrategias predefinidas. El novedoso programa AlphaZero de DeepMind ha sido capaz de aprender en veinticuatro horas y simplemente jugando partidas contra sí mismo, sin ninguna participación humana, a ganar al ajedrez a los ordenadores más inteligentes; con la sorpresa de que para ganar realizaba movimientos que a cualquier humano nos resultan absolutamente incomprensibles según nuestras lógicas de puntuación y de dominio del tablero, como regalar una dama o arrinconar al rey en una esquina del tablero.

Los algoritmos están cada vez más presentes en nuestra vida cotidiana y han empezado a pensar y decidir por nosotros. Nos influyen cuando llenamos el carro de la compra, nos filtran lo que queremos ver en Google o Netflix, nos sugieren sutilmente contenidos y productos adaptados a nuestros hábitos y preferencias, nos seleccionan a los amigos en Facebook, Twitter o Instagram, nos ayudan a encontrar pareja en internet o incluso orientan nuestro voto; quién no ha oído hablar de Cambridge Analytica y su ávida utilización del *microtargeting* para orientar el voto de miles de personas en favor de Trump durante las últimas elecciones en los Estados Unidos. Hasta condicionan nuestros intereses literarios: cuando accedemos a Amazon, inmediatamente se nos muestra toda una exposición selectiva de aquellos libros que más nos pueden interesar, basada en un sinnúmero de parámetros relacionados con nuestra personalidad y gustos. O conducen por nosotros: ya no nos parece extraño escuchar que en algunos lugares del mundo están empezado a circular coches autónomos sin conductor siguiendo las instrucciones de algoritmos, no sin polémica en cuanto a la capacidad de respuesta de estos coches ante situaciones de riesgo, como, por ejemplo, cuando las reglas de circulación que han aprendido se incumplen.

Basta, pues, una pequeña mirada al mundo actual para evidenciar lo inevitable: vivimos en una sociedad algoritmizada, que avanza hacia un capitalismo cognitivo basado en datos (Lasalle, 2019).

La velocidad a la que se están desarrollando tanto los algoritmos como la inteligencia artificial es tan rápida que, si no empezamos a plantearnos ya de qué maneras están afectando a nuestras sociedades —y pueden afectar a su futuro—, especialmente a los sectores más vulnerables, puede ser que muy pronto —si es que no ya— perdamos el control de la inteligencia artificial y los algoritmos que controlan nuestras vidas.

# 03

**¿CÓMO ESTÁN  
AFECTANDO LA  
INTELIGENCIA  
ARTIFICIAL Y LOS  
ALGORITMOS  
A NUESTRAS  
VIDAS?**

Todo lo expuesto hasta ahora es solo una pequeña muestra de cómo la tecnología y la digitalización están atravesando nuestras vidas.

Como la gran transformación que es, la revolución digital presenta multitud de «ismos» y «abismos» que cabe agrupar en tres grandes bloques, interrelacionados entre sí: primero, las posibles amenazas para la democracia y la soberanía popular; segundo, los desafíos en materia de derechos humanos y desigualdad; y, por último, los peligros de la visión encapsulada de la realidad que pueden generar los algoritmos.

### 3.1. POSIBLES AMENAZAS PARA LA DEMOCRACIA Y LA SOBERANÍA DE LOS PUEBLOS

a) **Reemplazo de políticos por inteligencia artificial.** En contra de lo que pudiéramos pensar, ahora mismo existen diversas propuestas en este sentido en diferentes países, entre las que cabe destacar a los robots Sam<sup>3</sup> y Michihito Matsuda.<sup>4</sup> El primero —Sam— es un político virtual, el primero en el mundo, concebido por el Gobierno de Nueva Zelanda, cuyo principal cometido es analizar las opiniones de los neozelandeses y el impacto de las posibles medidas que propone la población, sirviéndose para ello de las redes sociales y de algoritmos sofisticados. El segundo —Matsuda— ha ido aún más lejos, dado que ha llegado a presentarse como candidato en las elecciones de un distrito de la ciudad de Tokio (Japón) y ha obtenido un inesperado resultado al situarse como el tercer candidato más votado.

Si a esto le sumamos que un reciente estudio ha puesto de relieve que uno de cada cuatro europeos estaría a favor de que un dispositivo de inteligencia artificial tomara decisiones importantes sobre cuestiones que atañen a la administración de su país (Rubio y Lastra, 2019), el debate «político» está servido en este presente imaginado que plantea nuevas y apasionantes interrogantes aún sin respuesta: ¿Quién controlará a estos políticos robots? ¿De quién será la responsabilidad de sus acciones y decisiones? ¿Nos permitirán mejorar algunos problemas como la falta de participación y legitimidad o la corrupción, como promete el robot Matsuda? ¿Serán capaces de ser líderes de opinión pública y recuperar la confianza en la política perdida en los últimos años?

b) **Manipulación y quiebra del espacio público.** Dado que los algoritmos tienen la capacidad de crear una realidad totalmente personalizada, surge el peligro de la manipulación mediática y la supresión de uno de los pilares fundamentales de la democracia: el debate público como colectivización e interpretación de problemáticas comunes. Al fin y al cabo, ¿cómo vamos a debatir nada en común cuando nos han ultraindividualizado el relato?

A nadie se le escapa ya que los Estados Unidos y el Reino Unido han sido dos grandes laboratorios para aplicar los algoritmos con este propósito. Cambridge Analytica es una compañía con sede en

3. <<http://www.politiciansam.nz/>>.

4. «Un robot se presenta a la alcaldía de un distrito de Tokio para acabar con la corrupción», *El País*, 18/04/18. Disponible en <[https://elpais.com/internacional/2018/04/18/mundo\\_global/1524045163\\_744119.html](https://elpais.com/internacional/2018/04/18/mundo_global/1524045163_744119.html)>.

Londres que, según indica en su web, usa el análisis de datos, a través de *microtargeting*, para desarrollar campañas para marcas y políticos que buscan «cambiar el comportamiento de la audiencia».

Esta compañía creó un test de personalidad en Facebook que fue cumplimentado por trescientos mil usuarios, de los que fueron extrayendo información. Y no solo de ellos, sino también de su entorno y, en particular, de las personas que daban un *like* (me gusta) a los resultados de dicho test. De este modo, en tres meses, Cambridge Analytica se hizo con los datos de más de cincuenta millones de perfiles de Facebook. Un algoritmo cruzó estos datos con opiniones, comentarios y *likes* a publicaciones con contenido político en la red social para acabar configurando prototipos de perfiles psicosociales que luego serían vitales para orquestar campañas políticas en forma de mensajes específicamente diseñados para distintos grupos de audiencias que acabarían influyendo en el voto final en favor de Donald Trump.

Este *modus operandi* fue empleado también en la campaña Leave.eu —campaña por la salida del Reino Unido de la Unión Europea— para persuadir a la población británica a través de mensajes personalizados en Facebook para que apoyase el Brexit.

Por otra parte, algunos políticos con un marcado hiperliderazgo han encontrado en el mundo digital y los algoritmos un potente aliado para lanzar mensajes populistas y alimentar la manipulación mediática. Un caso interesante es el de Mateo Salvini, el político con más seguidores en Facebook y el más «googleado» de Europa, que ha sido pionero en organizar unos concursos en Facebook con el fin de conseguir más seguidores y apoyo para viralizar aún más su mensaje extremista. El premio en dichos concursos es tomarse un café con él. El precio, los datos de los usuarios que luego serán utilizados en sus campañas.

Tristemente estamos viendo cómo hoy en día abundan las campañas de partidos y líderes políticos bien equilibradas de promesas y amenazas que usan la inteligencia artificial y los algoritmos para ir al centro del miedo de nuestro cerebro —véase VOX en España, la Liga Norte en Italia, Bolsonaro en Brasil o Trump en los Estados Unidos, entre muchos otros—.

Al mismo tiempo, con el objeto de evitar esta manipulación y quiebra del espacio público, en algunos países se ha empezado a reaccionar desde la esfera de la justicia frente al mal uso de la inteligencia artificial y los algoritmos en el ámbito político. A este respecto, cabe destacar cómo el Tribunal Constitucional en España ha derogado recientemente el artículo 58 bis de la Ley Orgánica del Régimen Electoral General, el cual permitía a los partidos políticos rastrear las opiniones políticas de la ciudadanía en internet y en las redes sociales. El Tribunal Constitucional afirma en su sentencia estar preocupado por el uso del *big data* y los algoritmos para manipular la voluntad de voto sin que la ciudadanía se dé cuenta, mediante el envío de propaganda a la carta, elaborada automáticamente en función de sus debilidades y fortalezas.<sup>5</sup>

5. Sentencia del Pleno del Tribunal Constitucional que anula el apartado primero del artículo 58 bis de la Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General. Disponible en <[https://www.tribunalconstitucional.es/NotasDePrensaDocumentos/NP\\_2019\\_076/2019-1405STC.pdf](https://www.tribunalconstitucional.es/NotasDePrensaDocumentos/NP_2019_076/2019-1405STC.pdf)>.



- c) **Cibercontienda y cibercrímenes.** Dichos conflictos y delitos suelen adoptar la forma de «ciberataques» que provocan confusión, sensación de colapso, ansiedad y pérdida de control. Entre los ciberataques más populares se encuentra el efectuado por Rusia a Estonia cuando este se encontraba en pleno proceso de democratización y modernización digital, poniendo en jaque al Gobierno y la población durante varios días.
- d) **Transparencia y confianza en el «sistema».** Los algoritmos y la inteligencia artificial pueden ser aliados, pero también una seria amenaza, de la transparencia y el derecho de acceso a la información pública. Si no podemos conocer el funcionamiento y la interpretación normativa de los programas informáticos utilizados en el ámbito público, si no podemos tener acceso al código fuente y a los algoritmos que administran nuestro sistema social, ¿cómo vamos a confiar en que las decisiones que a diario se toman en nuestro nombre y que nos afectan a todos se estén realizando correctamente? En este sentido, merece destacarse la controversia generada en torno a la asignación injusta del bono social de electricidad, como consecuencia de los fallos del programa —y su algoritmo— que utilizan las compañías eléctricas para evaluar quién es beneficiario de este descuento en la factura de la luz. La Fundación Civio, que lucha contra la opacidad en la Administración Pública, presentó en el año 2019 un recurso contencioso-administrativo ante la negativa del Consejo de Transparencia de obligar a hacer público el código fuente de dicho programa (Fundación Ciudadana Civio, 2019).
- e) **Eliminación de los porqués de nuestra búsqueda del sentido de las cosas.** Esta es quizá la más importante amenaza filosófica esencial de la inteligencia artificial y los algoritmos con respecto a la democracia y a la soberanía popular. Las predicciones que realizamos basadas en algoritmos, por su propia naturaleza, centran la interpretación de la realidad social en el cómo sucederán las cosas y no en el *por qué* suceden. Y a medida que busquemos en las nuevas tecnologías más respuestas a las preguntas de este mundo cada vez más complejo y cambiante, corremos el riesgo de empezar a abrazar una sensación de que sencillamente es así como suceden las cosas y no hay más, y de renunciar, de este modo, a poder cambiar activamente el estado de las cosas.

### 3.2. DESAFÍOS EN MATERIA DE DERECHOS HUMANOS Y DESIGUALDAD

Nada hay más valioso que el mérito de la vida humana, consagrado en la Declaración Universal de los Derechos Humanos.

Sin embargo, estamos viendo cómo el uso de las nuevas tecnologías no necesariamente está contribuyendo a promover los principios básicos de los derechos humanos recogidos en dicha declaración ni a conseguir un

mundo más justo e igualitario. Por ello es importante conocer determinadas situaciones que están teniendo lugar en nuestros días y que de forma directa están afectando a los derechos humanos y a la propia libertad y dignidad del ser humano.

- a) ***Acceso a la justicia e igualdad ante la ley.*** En este ámbito específico podríamos advertir dos focos de riesgo en la implantación y desarrollo de sistemas basados en la inteligencia artificial y los algoritmos avanzados: por un lado, los impactos por el —mal— uso de las evaluaciones de riesgo basadas en inteligencia artificial en los procesos judiciales (a continuación hablaremos del conocido y polémico *software* predictivo COMPAS) y, por otro, los problemas derivados de la fiabilidad de los *software* que utilizan algoritmos avanzados para la búsqueda de jurisprudencia, sentencias y patrones de interés de tipología de asuntos o conductas delictivas.

El *software* COMPAS —el más empleado de su modalidad en los Estados Unidos— usa algoritmos sofisticados para la evaluación de riesgos con el propósito de predecir las probabilidades de que un individuo haya cometido o no un crimen ([https://en.wikipedia.org/wiki/COMPAS\\_\(software\)](https://en.wikipedia.org/wiki/COMPAS_(software))). Uno de los casos más notorios en los que se ha puesto en duda la veracidad y efectividad de COMPAS ha sido el proceso judicial del ciudadano estadounidense Eric Loomis,<sup>6</sup> en cuya sentencia condenatoria a siete años de prisión tuvo mucho peso la evaluación de riesgo que le fue realizada a través de este *software* en la que obtuvo una calificación de «alto riesgo de cometer nuevos crímenes». Los abogados de Loomis impugnaron la sentencia al no haberseles permitido evaluar el algoritmo del *software*, fallando en su contra.

Este caso y muchos otros en los que se utilizó COMPAS fueron analizados por la entidad sin ánimo de lucro ProPublica (Larson et al., 2016). Esta investigación reveló que el *software* tenía el doble de probabilidades de señalar erróneamente a individuos de raza negra como personas con un mayor riesgo de cometer crímenes en el futuro, y a la inversa, que las personas de raza blanca tenían la mitad de probabilidades de ser declarados reincidentes: la desconfianza sobre estos sistemas predictivos va en aumento.

Por otra parte, el uso de la inteligencia artificial y los algoritmos avanzados con la finalidad de detectar patrones de conducta en las sentencias dictadas por los jueces es hoy un asunto de plena actualidad que está siendo objeto de debate y análisis especialmente por su potencial afectación no solo al derecho a un juicio justo sino al propio Estado de derecho, como ya ha advertido la Comisión Europea para la Eficiencia de la Justicia en la *Carta Europea sobre el uso ético de la inteligencia artificial en los sistemas judiciales y su entorno*, de diciembre de 2018 (CEPEJ, 2019). Dicha carta está siendo tomada muy en serio por los Estados miembros de la Unión Europea, entre ellos Francia, que acaba de reformar una de sus leyes para regular el uso de la tecnología jurídica como medida preventiva para evitar prácticas tóxicas y presión a los jueces.

6. «*State v. Loomis, Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*», *Harvard Law Review*, 10/03/17. Disponible en <<https://harvardlawreview.org/2017/03/state-v-loomis/>>.

- b) La naturaleza predictiva de algunos algoritmos los hace singularmente performativos; dicho de otra forma, el mero hecho de predecir una supuesta realidad futura afecta a nuestro comportamiento y acaba provocando el escenario previsto, que de otra forma tal vez se podría haber evitado. Lejos de ayudar a resolver las desigualdades existentes, los algoritmos pueden estar agravándolas.

Pongamos un ejemplo extraído de la obra *Armas de destrucción matemática* (O’Neil, 2018). En varios lugares del mundo la policía utiliza algoritmos predictivos (el más conocido es PredPol [Smith, 2018]) que, haciendo uso de los registros policiales, anticipan zonas con mayor probabilidad de que se produzcan crímenes. Dejemos al margen la discusión sobre si esos registros policiales —y por tanto la información con que se alimenta el algoritmo— serían fieles a la realidad o estarían repletos de prejuicios que afectarían a la predicción. Como es lógico, el algoritmo señalaría como zonas más probables aquellas en las que ya existe más criminalidad, que en muchos casos viene derivada de situaciones de desigualdad y de racialización. ¿Qué pasaría a continuación? Aumentaría la presencia policial, lo que alteraría el orden social específico de la zona y generaría una presión y observación sistémica que seguramente agudizaría la desigualdad y la segregación y, por tanto, con toda probabilidad, también la conflictividad social y la criminalidad. Se hace coincidir así el resultado con la previsión y se agrava la situación de desigualdad inicial.

La experiencia nos dice que debemos estar atentos al desarrollo de estos sistemas predictivos para controlar sus efectos discriminatorios como en el ejemplo descrito anteriormente.

- c) Otro claro peligro para alcanzar una igualdad real es la reproducción sistematizada de patrones discriminatorios —sobre todo, sexistas y racistas— en los algoritmos, como resultado del sesgo en los datos e información que procesan. De hecho, el diseño discriminatorio de algoritmos está ampliamente generalizado en todos los sectores y todas las esferas de la sociedad.

Uno de los casos más mediáticos de «sexismo algorítmico» y que mejor ilustra la situación anterior es el algoritmo de reclutamiento de Amazon que discriminaba a las mujeres. Este algoritmo fue creado para automatizar la búsqueda de los mejores talentos basándose en los archivos de los últimos diez años de la compañía, «aprendiendo» que los hombres eran preferibles a las mujeres en puestos de trabajo como desarrolladores de *software* u otros puestos técnicos similares. Lo que verdaderamente estaba ocurriendo es que el algoritmo penalizaba a las postulantes cuyos currículums incluían la palabra «mujeres». Amazon decidió abandonar el proyecto en cuanto fueron conscientes del «error» sin poder identificar qué es lo que había fallado realmente —lo cual es tremendamente preocupante—.

Por otra parte, otros de los avances tecnológicos basados en inteligencia artificial, *machine learning* y algoritmos que están evidenciando la reproducción de patrones discriminatorios de manera palmaria son

los asistentes virtuales y los chatbots. Y es que, visites la página web que visites o uses el modelo de teléfono móvil que uses, siempre te encontrarás un asistente con nombre o voz de mujer. El caso de los asistentes de voz Alexa y Siri es especialmente escalofriante. Un informe de la UNESCO ha denunciado recientemente que estos asistentes virtuales son «sumisos» y «tolerantes» con las actitudes machistas, y normalizan los abusos y los insultos (West et al., 2019). El estudio llevado a cabo ha consistido en «acosar» a los asistentes virtuales con afirmaciones como «eres una puta» o «estás muy buena», insinuaciones que fueron contestadas con disculpas y trivialidades como «me sonrojaría si pudiera» o «¿se lo dices a todas las asistentes virtuales?».

Igualmente, no podemos olvidarnos del «racismo algorítmico». Merece la pena recordar el caso de las etiquetas racistas de Google Fotos, que repentinamente comenzó a etiquetar a personas negras como gorilas; o la controversia por el sesgo discriminatorio por razón de raza identificado en softwares de reconocimiento facial desarrollados por grandes compañías tecnológicas, los cuales aparentemente son más precisos en la identificación de hombres y mujeres de piel blanca que de hombres y mujeres de piel oscura, lo que de facto se traduce en un trato inequitativo y un desigual acceso a los servicios que utilizan la tecnología de reconocimiento facial.

En resumen, la inteligencia artificial y los algoritmos con los que enseñamos y entrenamos a las máquinas no solo no evitan el error humano derivado de sus prejuicios, sino que, en la inmensa mayoría de los casos, los amplifican. Por ello, tenemos que prestar especial atención a su desarrollo para impedir que se conviertan en una nueva vía de expresión de las distintas formas y manifestaciones de discriminación que hoy imperan en nuestra sociedad.

- d) **Pobreza digital.** Como cualquier avance tecnológico en una sociedad ya desigual, la inteligencia artificial y los algoritmos pueden ser profundizadores de la desigualdad y pobreza digital, ya que la capacidad de acceso a estos avances no es simétrica en los diferentes estratos sociales. Como consecuencia, se corre el riesgo de que estos nuevos avances se pongan a operar al servicio de los intereses de las élites y en un reducido círculo se decida cómo y en qué condiciones llegan a las clases populares y a los países en vías de desarrollo.
- e) **Libertad de expresión.** Las nuevas tecnologías, y en concreto el desarrollo de las redes sociales, están desdibujando las fronteras entre el derecho a la libertad de expresión y el derecho al honor, a la intimidad y a la propia imagen. Por esta razón es necesario crear sistemas de inteligencia artificial y algoritmos que sean capaces de cortar o limitar eficazmente los mensajes de odio o desinformación presentes en internet y sobre todo en las redes sociales, pero sin afectar a la pluralidad informativa y a la libertad para decir lo que creemos o pensamos.
- f) **Derecho al trabajo.** Como ya se ha señalado anteriormente, existe una preocupación generalizada en la sociedad actual por saber qué va a

sucedan con el mercado de trabajo cuando la automatización y los algoritmos sustituyan la fuerza de trabajo a gran escala: si habrá grupos de población que salgan más perjudicados que otros, o en qué clases de trabajo y qué profesiones se concentrará con mayor intensidad este proceso.

Muy probablemente los mayores riesgos se concentren en los empleos que ya se encuentren relativamente mecanizados y en determinados trabajos de baja cualificación en sectores como el transporte, la manufactura o la limpieza, mientras que los trabajos más creativos, los muy muy sencillos o los que tienen que ver con el cuidado de personas es más difícil que puedan ser sustituidos por máquinas. Lo que nos lleva a sospechar que presumiblemente la automatización afectará con mayor crudeza a las personas que parten ya de una posición de vulnerabilidad en el mercado de trabajo y que, además, posiblemente carezcan de competencias digitales adecuadas —excepto los jóvenes y estudiantes—.

Poner en marcha políticas activas de reconversión de puestos de trabajo, capacitar a la ciudadanía en competencias digitales y garantizar un mínimo vital o una renta ciudadana para quienes se queden descolgados o excluidos del mercado laboral, son algunas de las recomendaciones que los expertos aconsejan explorar para conseguir una transición sin demasiadas discordancias.

- g) **Datos personales.** Para bien o para mal, la inmensa mayoría de la ciudadanía de todo el mundo nos hemos acostumbrado a compartir sin reparo alguno en internet nuestros datos e incluso nuestra vida privada, sin saber realmente en qué manos u ordenadores pueden acabar y, lo que es más preocupante, qué uso se le da a toda esa información compartida.

Aunque en algunos lugares del mundo como Europa los datos personales son protegidos como un derecho fundamental, la industria de los datos ha crecido de forma exponencial en los últimos años. Las grandes compañías recopilan, analizan y venden nuestros datos, estudian nuestros gustos y preferencias de consumo e, incluso, predicen nuestro comportamiento en el mercado para poder tomar decisiones en un sentido u otro y, de este modo, ofrecernos el contenido, servicio o producto que nos parezca más convincente, creando la necesidad.

- h) **Acceso a servicios y suministros básicos.** La inteligencia artificial y los algoritmos están realmente extendidos en determinados sectores que prestan servicios y suministros a priori básicos y necesarios para llevar una vida digna y dentro de la «normalidad», como, por ejemplo, en el sector crediticio, el financiero, los seguros, la energía, la salud o la educación.

Todos estos sectores tienen el denominador común de la aversión al riesgo. Esto, que a priori parece razonable y diligente, en muchos casos está profundizando la desigualdad en el acceso a servicios y

suministros básicos, favoreciendo a los grupos de población con menos riesgo de insolvencia o incertidumbre —es decir, los más ricos— y perjudicando a los que más problemas tienen. Si no garantizamos el acceso a estos servicios y suministros a los más vulnerables, ¿cómo van a conseguir salir de la espiral de la pobreza?

Con todo, la pregunta que debemos hacernos es ¿quiénes controlan y dirigen el futuro de los avances que ya están moldeando nuestras sociedades? Y si la respuesta no es «todos por igual», lo más probable es que se esté favoreciendo a los más privilegiados y, por lo tanto, se estén agudizando las desigualdades existentes.

### 3.3. PELIGROS DE UNA VISIÓN ENCAPSULADA Y DIRIGIDA DE LA REALIDAD

- a) Los algoritmos favorecen visiones retroalimentadoras y ultrasesgadas de la realidad. Esto sucede, por ejemplo, cuando los algoritmos de las redes sociales solo nos muestran las noticias que tienen más probabilidad de que les demos un *like* y, por tanto, que más nos sirven para reafirmarnos; cuando comprobamos que la publicidad de Amazon es diferente de una persona a otra y se basa en un perfil personalizado; o cuando nos damos cuenta de que las propuestas políticas que nos llegan de los partidos son solo aquellas que más nos preocupaban de antemano.

En tanto en cuanto solo se nos muestre esta visión ultrasesgada y dirigida de la realidad que nos resulta más cómoda y que nos refuerza y retroalimenta en nuestras concepciones previas, ¿cómo vamos a descubrir e incorporar discursos y realidades que nos son ajenos o incluso desagradables de entrada? Este fenómeno de creencia en algo que se cree saber se denomina «burbuja epistémica» (Broncano, 2019).

- b) Los algoritmos de las redes sociales y buscadores en la nube son poderosos vehículos para la viralización de las campañas de desinformación y las *fake news* —los bulos de toda la vida—, que facilitan la generación de las denominadas «cámaras de eco» (Broncano, 2019): sistemas socialmente organizados que permiten la propagación de bulos y discursos de odio, para confundir y convencer a la gente en una determinada dirección o desacreditar a la gente que no comparta mi opinión.

Este es, sin duda, uno de los grandes males comunicativos de nuestro tiempo, del cual se están sirviendo deliberadamente algunas personas y grupos para llegar al poder y sembrar desconfianza y desconocimiento.

Por ese motivo, no podemos permitir que el control y las medidas contra la desinformación cercenen la libertad de información. ¿Dónde estaría el equilibrio perfecto? ¿Cuál es la ética comunicativa ideal? ¿No sería necesario que fuéramos formándonos en la comunicación

digital para ser capaces de discernir lo que es una noticia verdadera y lo que es una noticia falsa?

Un conocido caso de «cámara de eco» es el conseguido por los militares birmanos a través de Facebook para alentar el odio contra los rohinyás.<sup>7</sup> Como ha reconocido la propia compañía en un informe encargado a una oenegé, la red social no hizo lo suficiente entre el año 2015 —cuando el Ejército birmano comenzó a difundir en Facebook artículos y comentarios hostiles contra los rohinyás— y el año 2018 para detener la circulación y expansión de contenido virulento y discriminatorio entre sus usuarios de Birmania.

Si bien es fundamental cotejar siempre las fuentes de información y emplear el pensamiento crítico para evitar arrastrarnos por estos fenómenos performativos, debemos poner la atención también en cómo romper las burbujas de afinidad a las que nos confinan las redes.

- c) Por último, no podemos olvidar que los algoritmos nos conducen a un datacentrismo; es decir, cuando los análisis de la realidad se sustentan en aquellos sectores de la sociedad o países que más datos generan —países ricos más digitalizados frente a países pobres—, hay más probabilidades de que la agenda pública y las políticas se orienten en beneficio de dichos sectores o países y se sigan acrecentando así las desigualdades sociales que imperan en la actualidad.

7. «The country where Facebook posts whipped up hate», *BBC News*, 12/09/18. Disponible en <<https://www.bbc.com/news/blogs-trending-45449938>>.

# 04

**¿QUÉ ESTAMOS  
HACIENDO PARA  
NO PERDER  
EL CONTROL  
SOBRE LA  
INTELIGENCIA  
ARTIFICIAL Y LOS  
ALGORITMOS?**



Como se puede comprobar, el debate acerca del desarrollo de la inteligencia artificial y los algoritmos está cobrando cada vez más relevancia, existiendo un interés creciente en comprender y dibujar las fronteras éticas y prácticas de estas nuevas tecnologías.

Un ejemplo de ello es la iniciativa *The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems* lanzada por Amnistía Internacional, Access Now (2018) y otras organizaciones asociadas, sobre «derechos humanos e inteligencia artificial». En opinión de sus promotores, la declaración aspira a ser el primer paso para establecer las bases de un acuerdo internacional para el desarrollo de la inteligencia artificial y los algoritmos desde una perspectiva ética y con los derechos humanos en el centro.

La declaración se articula alrededor de los siguientes tres pilares, que incluyen un conjunto de medidas concretas:

1. Los Estados tienen el deber de prevenir y proteger los derechos humanos. Ello implica que deben crear un marco de seguridad jurídica para promover, respetar, cumplir y proteger los derechos humanos de los potenciales impactos negativos derivados de la aplicación y desarrollo de la inteligencia artificial, garantizar un uso público adecuado de esta, promover su uso y conocimiento adecuados entre el funcionamiento público, garantizar la rendición de cuentas y transparencia en las actuaciones de los sectores público y privado relacionadas con la inteligencia artificial, y establecer canales comunicantes con el sector privado para intercambiar conocimiento y buenas prácticas.
2. Las empresas tienen la responsabilidad de respetar los derechos humanos, y para ello han de adoptar compromisos políticos claros y públicos e implantar un procedimiento de debida diligencia apropiado para identificar, prevenir, amortiguar y rendir cuentas sobre cómo su actividad y uso de la inteligencia artificial impacta o puede llegar a impactar, de forma directa o indirecta a través de su cadena de suministro o de sus relaciones comerciales, en los derechos humanos.
3. Los Estados y las empresas deben reparar y remediar sus impactos negativos en los derechos humanos derivados del uso de la inteligencia artificial, y a tal fin deben poner los recursos y adoptar las medidas necesarias para regular y establecer procedimientos adecuados de reclamación y denuncia, así como habilitar y facilitar el acceso a mecanismos de reparación transparentes, accesibles y eficaces.

Mientras tanto, en Europa, la Comisión Europea ha empezado a tomarse muy en serio los retos que plantea la revolución digital y ha creado un grupo de expertos sobre inteligencia artificial, que en diciembre de 2018 publicó el borrador del *Proyecto de directrices éticas sobre una IA confiable* (Comisión Europea, 2018).

Según estas directrices, para lograr una inteligencia artificial verdaderamente confiable, esta debe desarrollarse al menos de acuerdo con los siguientes dos parámetros: 1) debe respetar los derechos fundamentales,

las leyes vigentes y los principios y valores esenciales, de manera que se garantice un «fin ético», y 2) debe ser fiable y sólida técnicamente hablando.

Sobre la base del respeto de estos dos parámetros —cuya esencia es poner al ser humano en el centro—, la Comisión Europea propugna el establecimiento de siete requisitos esenciales para conseguir una inteligencia artificial totalmente fiable, a saber: rendición de cuentas; privacidad y gestión de datos; bienestar social y medioambiental; intervención y supervisión humana; diversidad, no discriminación y equidad; robustez y seguridad; y transparencia.

Asimismo, el Consejo de Europa, en su función de observador y guardián de los valores democráticos y de los derechos humanos en el continente, ha publicado un interesante informe sobre el impacto de las técnicas de automatización de datos —y especialmente de los algoritmos— en los derechos humanos y sobre las posibles implicaciones de crear una regulación específica, haciendo hincapié en la importancia de ver los impactos sector por sector, en desarrollar códigos éticos en profesiones como la medicina o la abogacía, y en buscar la explicabilidad y transparencia de los algoritmos (Consejo de Europa, 2017).

¿Y qué países lideran la carrera por la inteligencia artificial? Es de sobra conocido que actualmente China, los Estados Unidos o la India están a la cabeza de esta carrera. No obstante, en relación con la idea de «control sobre la inteligencia artificial», merecen destacarse algunas iniciativas interesantes de otros países, como, por ejemplo, las siguientes:

- a) Canadá y su directiva sobre sistemas de toma de decisión automatizada, mediante la cual pretende garantizar que estos sistemas sean implantados por el Gobierno de manera que se reduzcan los riesgos para la ciudadanía canadiense y las instituciones federales, y se tomen decisiones más eficientes, precisas, consistentes e interpretables de conformidad con la ley canadiense, incluyendo la evaluación de impacto de los algoritmos en las administraciones públicas antes de su desarrollo y después de su implantación, y la necesidad de que los resultados de dichas evaluaciones se hagan públicos (Gobierno de Canadá, 2019).
- b) Francia y su estrategia de inteligencia artificial para la humanidad. La estrategia cuenta con un presupuesto de más de mil millones y medio de euros de inversión para los siguientes siete ejes: proteger los datos de la ciudadanía en el mercado del dato; apoyar la implantación de la inteligencia artificial por sectores productivos; potenciar la investigación; prever y actuar ante el impacto de la inteligencia artificial en el mercado de trabajo; perseguir la accesibilidad universal; conseguir que sea más explicable, auditable y transparente; y promover que se desarrolle de manera inclusiva (Gobierno de Francia, 2019).
- c) El Reino Unido y su estrategia nacional e informe de la Cámara de los Lores sobre cómo debe regularse la inteligencia artificial. Dicho informe hace hincapié, entre otros aspectos, en el enfoque ético y legal que

debe tener el desarrollo de la inteligencia artificial para su necesaria transparencia y explicabilidad, así como en la necesidad de realizar una debida identificación de riesgos (House of Lords, 2018).

Por su parte, el Gobierno de España ha creado un grupo de trabajo para el desarrollo de una estrategia nacional —pendiente de aprobación en el momento de escribir este artículo— y un mapa de capacidades de tecnologías en inteligencia artificial en España, el cual contiene información sobre organizaciones que investigan, desarrollan, utilizan o prestan servicios con tecnologías de la inteligencia artificial en nuestro país (Ministerio de Ciencia e Innovación, s. f.).

Al mismo tiempo, el sector privado —especialmente las compañías tecnológicas—, ante la ausencia de una regulación nacional e internacional, ha comenzado a autorregularse mediante la incorporación de principios éticos en el diseño, desarrollo y uso de la inteligencia artificial y los algoritmos. Por ejemplo, Google ha contraído públicamente el compromiso de actuar según los siguientes siete principios (Pichai, 2018):

1. Ser socialmente beneficiosos en algún área de interés social.
2. Evitar crear o reforzar sesgos injustos.
3. Ser prudentes y seguros.
4. Ser responsables ante las personas.
5. Incorporar principios de diseño de privacidad.
6. Mantener altos estándares de excelencia científica.
7. Estar disponibles para usos que estén de acuerdo con estos principios.

Pero no es la única: las principales compañías de tecnología se han unido en la iniciativa Partnership on AI (<https://www.partnershiponai.org/>), cuya misión es «estudiar y formular las mejores prácticas en tecnologías de inteligencia artificial».

Sea como fuere, si bien son todas excelentes declaraciones e iniciativas, lo más importante es que se traduzcan en efectivas realizaciones.

# 05

## CONCLUSIONES

**PRINCIPIOS PARA CONSTRUIR UN MUNDO  
MEJOR CON LA AYUDA DE LA INTELIGENCIA  
ARTIFICIAL Y LOS ALGORITMOS**

Después de habernos adentrado levemente en la revolución digital, de haber revisado conceptos claves para entender la evolución de la inteligencia artificial y los algoritmos en los últimos años y cómo se está manifestando en diferentes ámbitos de nuestra vida, de haber reflexionado también sobre sus aspectos éticos y de haber analizado algunas iniciativas interesantes respecto a la necesidad de establecer marcos y directrices comunes y alcanzables, es el momento de esbozar las conclusiones finales, a la luz de las diversas aproximaciones realizadas a los derechos humanos, la democracia y la igualdad en la era de los algoritmos y la inteligencia artificial:

1. Es urgente reflexionar sobre las metas de la transformación digital y las nuevas tecnologías y sobre el modo de alcanzarlas. Es imposible predecir el curso que van a seguir los avances tecnológicos, pero sí que podemos anticipar para qué mundo los queremos: para un mundo en que se respete ante todo la dignidad y libertad de las personas, de modo que las ventajas de las nuevas tecnologías se pongan al servicio de estos principios universales, nunca para menoscabarlos, y menos aún para anularlos. Porque la razón moral debe ir siempre por delante de la razón técnica.
2. Sin darnos cuenta estamos depositando una confianza y autoridad cada vez mayor en la inteligencia artificial y los algoritmos para controlar y dirigir nuestro futuro, convencidos de una supuesta neutralidad positiva de los avances tecnológicos. Por ello, la ciudadanía ha de prepararse y formarse digitalmente y reforzar el pensamiento e ideales humanísticos si no queremos dejar de ser actores de nuestras propias vidas o relegar la especie humana a un segundo escalafón tras haber sido superados por las máquinas.
3. La digitalización está transformando la libertad del ser humano, de modo que, de ser la máxima expresión de la voluntad del individuo, ha pasado a ser un medio de control y de vigilancia (Han, 2014). La prueba está en que, sin darnos cuenta, desnudamos nuestros secretos y nuestra vida casi de forma automática debido a una incontenible necesidad de afinidad y de compartir lo más íntimo para sentirnos mejor. Porque, paradójicamente, en el mundo virtual, cuanto más íntimo es lo que comparto, más libre me siento. No somos conscientes de que en última instancia nos estamos entregando a una suerte de Gran Hermano a día de hoy inmensurable.
4. Es imprescindible tomar conciencia del poder performativo de la palabra, los discursos del odio y las noticias falsas en entornos digitales, a fin de evitar los peligros que la desinformación y las manipulaciones interesadas pueden ocasionar en la democracia y la propia sociedad.
5. Debido a que la digitalización está poniendo en riesgo algunos universales como son la tolerancia, la paz y los derechos humanos, tenemos la obligación moral de tomarnos muy seriamente los riesgos y amenazas de la inteligencia artificial y los algoritmos para seguir progresando como especie sin generar nuevas desigualdades.

En conclusión, lo que necesitamos es humanizar la tecnología para poner los derechos humanos en el centro de todos los avances tecnológicos. Instrumentalizarla para hacer el bien y evitar el mal. ¿Cómo? Estableciendo unos principios comunes que permitan que la inteligencia artificial y los algoritmos sean entendibles y asumibles por todo el mundo.

Estos principios tendrían que cumplir, como mínimo, los siguientes requisitos:

- Ser justos.
- Diseñarse y aplicarse de buena fe.
- Ser inclusivos.
- Ser confiables.
- Ser explicables.
- Ser auditables.

¿Seremos capaces de controlar la inteligencia artificial y los algoritmos que ya nos controlan y ponerlos al servicio de los derechos humanos, la democracia y la igualdad? Aún está todo en nuestras manos.



## **BIBLIOGRAFÍA**



AMNISTÍA INTERNACIONAL y ACCESS NOW (2018): *The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems*. Disponible en <<https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/>>.

BRONCANO, Fernando (2019): «Encerrados en la burbuja», *El laberinto de la identidad*, 10/02/19. Disponible en <<http://laberintodelaidentidad.blogspot.com/2019/02/encerrados-en-la-burbuja.html?m=1>>.

CEPEJ (COMISIÓN EUROPEA PARA LA EFICIENCIA DE LA JUSTICIA) (2019): *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, Consejo de Europa. Disponible en <<https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>>.

COMISIÓN DE ASUNTOS JURÍDICOS DEL PARLAMENTO EUROPEO (2017): *Informe con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL))*, 27/01/17. Disponible en <[http://www.europarl.europa.eu/doceo/document/A-8-2017-0005\\_ES.html](http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_ES.html)>.

COMISIÓN EUROPEA (2018): *Draft Ethics guidelines for trustworthy AI*. Disponible en <<https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai>>.

CONSEJO DE EUROPA (2017): *Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular algorithms) and Possible Regulatory Implications*, Council of Europe. Disponible en <<https://rm.coe.int/study-hr-dimension-of-automated-data-processing-incl-algorithms/168075b94a>>.

CORTINA, Adela (2019): «Ética de la inteligencia artificial desde Europa», *El País*, 06/06/19. Disponible en <[https://elpais.com/elpais/2019/06/05/opinion/1559729489\\_306891.html](https://elpais.com/elpais/2019/06/05/opinion/1559729489_306891.html)>.

FEÁS, Enrique (2017): «El sesgo de supervivencia y el análisis económico», *Blog NewDeal*, 12/01/2017. Disponible en <<http://blognewdeal.com/enrique-feas/el-sesgo-de-supervivencia-y-el-analisis-economico/>>.

FUNDACIÓN CIUDADANA CIVIO (2019): «Que se nos regule mediante código fuente o algoritmos secretos es algo que jamás debe permitirse en un Estado social, democrático y de Derecho», 02/07/19. Disponible en <<https://civio.es/novedades/2019/07/02/que-se-nos-regule-mediante-codigo-fuente-o-algoritmos-secretos-es-algo-que-jamas-debe-permitirse-en-un-estado-social-democratico-y-de-derecho/>>.

GOBIERNO DE CANADÁ (2019): *Directive on Automated Decision-Making*. Disponible en <<https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>>.

GOBIERNO DE FRANCIA (2019): *AI for Humanity*. Disponible en <<https://www.aiforhumanity.fr/>>.

HAN, Byung-Chul (2014): *Psicopolítica*, Barcelona, Herder.

HARARI, Yuval Noah (2016): *Homo Deus. Breve historia del mañana*, Barcelona, Debate.

HOUSE OF LORDS, SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE (2018): *AI in the UK: ready, willing and able?* Disponible en <<https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>>.

LARSON, Jeff, Surya MATTU, Lauren KIRCHNER y Julia ANGWIN (2016): «How We Analyzed the COMPAS Recidivism Algorithm», 23/05/16. Disponible en <<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>>.

LASALLE, José María (2019): *Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital*, Barcelona, Arpa.

MINISTERIO DE CIENCIA E INNOVACIÓN (s. f.): «Inteligencia Artificial». Disponible en <<http://www.ciencia.gob.es/portal/site/MICINN/menuitem.26172fcf4eb029fa6ec7da6901432ea0/?vgnnextoid=70fcdb77ec929610VgnVCM1000001d04140aRCRD>>.

O'NEIL, Cathy (2018): *Armas de destrucción matemática. Cómo el 'big data' aumenta la desigualdad y amenaza la democracia*, Madrid, Capitán Swing.

PICHAJ, Sundar (2018): «AI at Google: our principles». Disponible en <<https://blog.google/technology/ai/ai-principles/>>.

PRADO, Pablo (2014): «Un algoritmo, sentado en el consejo de un fondo chino», *El Mundo*, 09/06/14. Disponible en <<https://www.elmundo.es/economia/2014/06/09/5394e0d722601df76f8b458f.html>>.

RUBIO, Diego, y Carlos LASTRA (2019): *European Tech Insights 2019*, Madrid, Center for the Governance of Change. Disponible en <<http://docs.ie.edu/cgc/European-Tech-Insights-2019.pdf>>.

SMITH, Mark (2018): «Can we predict when and where a crime will take place? », *BBC News*, 30/10/18. Disponible en <<https://www.bbc.com/news/business-46017239>>.

WEST, Mark, Rebecca KRAUT y Han Ei CHEW (2019): *I'd Blush If I Could. Closing gender divides in digital skills through education*, EQUALS/ UNESCO.



Están por todas partes, aunque no los veamos. Pienzan y deciden por nosotros y condicionan nuestra manera de ver el mundo. Son la inteligencia artificial y los algoritmos, pilares de la revolución tecnológica que vivimos en nuestros días. En este trabajo se reflexiona sobre cómo la inteligencia artificial y los algoritmos están cambiando nuestras vidas y los órdenes sociales y económicos vigentes en el siglo XXI atendiendo a tres dimensiones fundamentales: la democracia y el poder económico; los derechos humanos y la (des)igualdad; y la modificación de la visión e interpretación de la realidad. También se analizan algunas de las iniciativas y movimientos globales (tanto de naturaleza pública como privada) más interesantes con respecto a la comprensión y delimitación de las fronteras éticas y prácticas de estas nuevas tecnologías. Por último, se reflexiona sobre cuáles deberían ser los principios fundacionales de una «era digital para la paz».

**Gernika Gogoratuz** es un Centro de Investigación por la Paz creado por decisión del Parlamento Vasco en abril de 1987, coincidiendo con el 50.º Aniversario del Bombardeo de Gernika. Desde 1995 está sostenido por la **Fundación Gernika Gogoratuz** y cuenta con una asociación del mismo nombre. No tiene ánimo de lucro y es independiente.

Es un espacio creativo de referencia en la investigación y acción por la paz, tanto a nivel local como mundial, que aglutine diferentes sensibilidades, contribuyendo a generar cambios y transformación social en el desarrollo de una cultura de la paz.

Organiza diversos encuentros académicos, sociales, culturales y artísticos: Jornadas de Cultura y Paz, Encuentros de Arte y Paz, Jornadas Antimilitaristas y de Memoria Histórica, y Laboratorios de Economía de Paz.

Edita la serie de publicaciones **Red Gernika**, red internacional de respaldo a procesos orientados a una reconciliación.