

138  
diciembre  
2024

afers  
revista crítica d'  
internacionals

# Desigualdad algorítmica

## Gobernanza, representación y derechos en la IA

**Carme Colomina Saló, Daniel Innerarity  
y Marta Cantero Gamito (Coord.)**

# La cultura pasa por aquí



**arce**

ASOCIACIÓN  
DE REVISTAS  
CULTURALES  
DE ESPAÑA

C/ Orfila, 3. 2º Izda. 28010 Madrid (España) | Tel.: 34 91 308 60 66 | E-mail: secretaria@arce.es

[www.revistas culturales.com](http://www.revistas culturales.com) | [www.quioscultural.com](http://www.quioscultural.com) | [www.arce.es](http://www.arce.es)



ARCE, Revistas  
Culturales de España



arcerevistas culturales



RC\_ARCE



El Sillón Rojo

CON LA COLABORACIÓN DE

**CEDRO**

CENTRO ESPAÑOL DE DERECHOS REPROGRÁFICOS

DESIGUALDAD  
ALGORÍTMICA  
Gobernanza, representación  
y derechos en la IA **138**  
Diciembre 2024

Coordinación científica:  
Carme Colomina Saló, Daniel Innerarity  
y Marta Cantero Gamito

Proyecto editorial cofinanciado por la Unión Europea (UE). Las opiniones expresadas en los artículos son responsabilidad exclusiva de sus autores/as y no reflejan necesariamente los puntos de vista de la UE o la Agencia Ejecutiva Europea de Educación y Cultura (EACEA).



Cofinanciado por  
la Unión Europea

Director/*Editor-in-chief*: Pol Morillas  
Editora/*Managing Editor*: Elisabet Mañé  
Editora de sección/*Section Editor*: Isabel Verdet

  
revista cidob d'  
**afers**  
internacionales

**Consejo editorial/*Editorial Board*:**

Anna Ayuso (CIDOB), Clàudia Canals (AVANCSA, Generalitat de Catalunya), Oriol Costa (UAB), Blanca Garcés (CIDOB), Robert Kissack (IBEI), Marga León (UAB), Salvador Martí Puig (UdG), Jordi Vaquer (Metropolis).

**Consejo asesor/*Advisory Board*:** Sergio Aguayo, El Colegio de México, A.C.; Manuel Alcántara, Universidad de Salamanca; José Antonio Alonso, Universidad Complutense de Madrid, Universidad de Columbia; Esther Barbé Izuel, Universitat Autònoma de Barcelona; Adrián Bonilla, Facultad Latinoamericana de Ciencias Sociales (FLACSO); Alison Brysk, University of California, Santa Barbara; Miguel Ángel Centeno, Princeton University; Noe Cornago, Universidad del País Vasco; Rafael Fernández de Castro, Center for U.S.-Mexican Studies, UC San Diego; Caterina Garcia Segura, Universitat Pompeu Fabra; Jean Grugel, University of Sheffield; Fernando Guirao, Universitat Pompeu Fabra; Daniel Innerarity, Instituto de Gobernanza Democrática (GlobeMance); Jacint Jordana, Institut Barcelona d'Estudis Internacionals (IBEI); Gemma Marín Muñoz, Universidad Autónoma de Madrid; Diego Muro, University of St Andrews; Ludolfo Paramio (DEP), Instituto de Políticas y Bienes Públicos, CSIC; José Antonio Sanahuja, Universidad Complutense de Madrid; Maria Regina Soares da Lima, Universidad Estatal de Rio de Janeiro; Max Spoor, International Institute of Social Studies of Erasmus University Rotterdam; Juan Gabriel Tokatlian, Universidad Di Tella, Buenos Aires; Fidel Tubino, Pontificia Universidad Católica del Perú; Pere Vilanova i Trias, Universitat de Barcelona.

**CIDOB**

Elisabets 12, 08001 Barcelona, Spain  
T. (34) 93 302 64 95 / F. (34) 93 302 21 18  
publicaciones@cidob.org / www.cidob.org

Precio de este ejemplar: 11€

Suscripción anual: España: 31€ (Europa: 39€) (Resto países: 44€)

Impresión: QP Print Global Services

ISBN: 978-84-18977-27-5 • ISSN: 1133-6595 • E-ISSN 2013-035X

DOI: doi.org/10.24241/rcia

Dep. Legal: B. 17.645-1983

Distribuye: Edicions Bellaterra, C. de la Foneria, 5-7, baixos. Manresa (Barcelona)  
https://www.bellaterra.coop/es

Diseño y maquetación: Joan Antoni Balcells

Revisión por expertos: Isabel Verdet

Web y soporte técnico: Sílvia Serrano

Suscripciones: Héctor Pérez. Envíos: Marta Lizana

Esta revista es miembro de:



Esta revista ha recibido una ayuda a la edición,  
del Ministerio de Cultura-Dirección General del Libro,  
del Cómic y de la Lectura



Indexada en:  
**Scopus**<sup>®</sup>



Certificada por la FECYT:



Patrons de CIDOB



Los artículos expresan las opiniones de los autores.

“Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra sólo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Dirijase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org), si necesita fotocopiar o escanear algún fragmento de esta obra”.

138  
Diciembre 2024

ISSN: 1133-6595  
ISBN: 978-84-18977-27-5  
www.cidob.org

© CIDOB (Barcelona Centre for International Affairs)

Creada en 1982, *Revista CIDOB d'afers internacionales* es una publicación cultural/académica cuatrimestral de relaciones internacionales y desarrollo. Pionera en el ámbito hispanohablante, ofrece al lector un análisis en profundidad de los temas internacionales desde diferentes puntos de vista y perspectivas, combinando información y análisis. La publicación está dirigida a la comunidad académica y al público interesado y/o implicado en general. Se edita en formato impreso y digital.

Los artículos publicados pasan por un proceso de evaluación externa por pares de anonimato doble y están indexados y resumidos en las siguientes bases de datos:

**Academic Search Complete-EBSCO**  
**CAHRUS Plus+** (Base de datos de revistas científicas de los ámbitos de las Ciencias Sociales y las Humanidades)  
**Dialnet** (Portal de difusión de la producción científica hispana)  
**DICE** (Difusión y Calidad Editorial de las Revistas Españolas de Humanidades y Ciencias Sociales y Jurídicas)  
**DOAJ** (Directory of Open Access Journals)  
**DULCINEA** (Derechos de copyright y las condiciones de auto-archivo de revistas científicas españolas)  
**ERIH PLUS** (European Reference Index for the Humanities and the Social Sciences)  
**ESCI-Web of Science** (Clarivate Analytics)  
**IPSA** (International Political Science Abstracts)  
**ISOC-Ciencias Sociales y Humanidades** (CSIC)  
**Latindex** (Sistema regional de información en línea sobre revistas científicas de América Latina, el Caribe, España y Portugal)  
**MIAR** (Matriu d'informació per a l'Avaluació de Revistes)  
**PIO** (Periodicals Index Online)  
**REDALYC** (Red de Revistas Científicas de América Latina y el Caribe, España y Portugal)  
**REDIB** (Red Iberoamericana de Innovación y Conocimiento Científico)  
**RESH** (Revistas Españolas de Ciencias Sociales y Humanas)  
**Scopus**  
**ULRICH'S** (Global serials directory)

**Desigualdad algorítmica: gobernanza, representación y derechos en la IA**

Introducción .....	7
<b>Daniel Innerarity</b> .....	11
Defensa y crítica de la gobernanza algorítmica	
<b>Carne Colomina y Marta Galceran</b> .....	27
Las <i>otras</i> geopolíticas de la IA	
<b>Marta Cantero Gamito</b> .....	51
Acceso a la justicia en tiempos de IA: ¿hacia una justicia <i>low-cost</i> ?	
<b>Carlos Saura García y Patrici Calvo</b> .....	73
Opinión pública masiva: colonización algorítmica y <i>shintetificación</i> de la esfera pública	
<b>Águeda Gómez Suárez y Rosa Mª Verdugo Matés</b> .....	97
Nuevas formas de poder digital en la red: dimensiones de la política sexual de las <i>Big Tech</i>	
<b>Anna Levitina</b> .....	121
Los seres humanos en la toma de decisiones automatizada en el marco del RGPD y la Ley de IA	
<b>Miguel Guillén Burguillos y Eloi Serrano Robles</b> .....	145
Oportunidades desde la economía social en la lucha contra las amenazas de la IA	
<b>Encarnación La Spina</b> .....	171
La regulación europea de la IA ante los sesgos y riesgos de discriminación algorítmica en contextos migratorios	
<b>José Luís Blasco Ejarque y Francisco Javier Tirado</b> .....	195
La llegada de la IA a las ciudades: urbanismo de plataformas y nuevos regímenes de gubernamentalidad	
<b>Reseñas de libros (temas)</b>	
<b>Javier Borrás</b> .....	219
La colonización algorítmica de la democracia	
<b>Candela Ollé</b> .....	221
Robots como reflejos de humanidad	
<b>Ronald Sáenz Leandro</b> .....	224
Datificación, gobernanza y poder algorítmico: impensar las IA desde América Latina	
<b>José Antonio Sanahuja</b> .....	226
La «guerra contra el género» como contestación del orden internacional	
<b>Allan David Rodríguez Aristizábal y Juan Camilo Restrepo Restrepo</b> .....	229
El conocimiento y la democracia como fuente de la legitimidad	

revista cidob d'

# afers

internacionals



## ¡Suscríbete!

RECIBIRÁS 3 NÚMEROS AL AÑO  
EN LA DIRECCIÓN QUE NOS  
INDIQUES

TARIFAS: 31 € ESPAÑA  
39 € EUROPA  
44 € OTROS PAÍSES

CONTACTA CON:

**[publicaciones@cidob.org](mailto:publicaciones@cidob.org)**

### PRÓXIMOS NÚMEROS

- 139 El impacto del Sur Global en el sistema internacional
- 140 La politización de la inmigración en las dos orillas del Mediterráneo: una perspectiva comparada
- 141 Asia en transición: más allá de la competición geopolítica

# CONTENTS

## **Algorithmic inequality: governance, representation and rights in AI**

Introduction .....	7
<b>Daniel Innerarity</b> .....	11
Defence and criticism of algorithmic governance	
<b>Carme Colomina and Marta Galceran</b> .....	27
The <i>other</i> geopolitics of AI	
<b>Marta Cantero Gamito</b> .....	51
Access to justice in times of AI: towards a low-cost justice?	
<b>Carlos Saura García and Patrici Calvo</b> .....	73
Mass public opinion: algorithmic colonisation and <i>synthetisation</i> of the public sphere	
<b>Águeda Gómez Suárez and Rosa M<sup>a</sup> Verdugo Matés</b> .....	97
New forms of digital power on the web: dimensions of the sexual politics of Big Tech	
<b>Anna Levitina</b> .....	121
Humans in automated decision-making under the GDPR and AI Act	
<b>Miguel Guillén Burguillos and Eloi Serrano Robles</b> .....	145
Opportunities from the social economy in the fight against the threats of AI	
<b>Encarnación La Spina</b> .....	171
The European regulation on AI in the face of the biases and risks of algorithmic discrimination in migration contexts	
<b>José Luís Blasco Ejarque and Francisco Javier Tirado</b> .....	195
The arrival of AI to the cities: platform urbanism and new regimes of governmentality	

### **Book reviews (subjects)**

<b>Javier Borrás</b> .....	219
The algorithmic colonisation of democracy	
<b>Candela Ollé</b> .....	221
Robots as reflections of humanity	
<b>Ronald Sáenz Leandro</b> .....	224
Datafication, governance and algorithmic power: unthinking AI from Latin America	
<b>José Antonio Sanahuja</b> .....	226
The "war on gender" as protest against the international order	
<b>Allan David Rodríguez Aristizábal and Juan Camilo Restrepo Restrepo</b> .....	229
Knowledge and democracy as a source of legitimacy	



# Introducción. Desigualdad algorítmica: gobernanza, representación y derechos en la IA

## Introduction. Algorithmic inequality: governance, representation and rights in AI

### Daniel Innerarity

Catedrático de Filosofía Política, investigador Ikerbasque, Universidad del País Vasco; titular de la cátedra de Inteligencia Artificial y Democracia, Instituto Europeo de Florencia. *dinner@ikerbasque.org*. ORCID: <https://orcid.org/0000-0003-4307-8468>

### Carme Colomina Saló

Investigadora sénior y editora, CIDOB; profesora asociada, College of Europe (Bélgica). *ccolomina@cidob.org*. ORCID: <https://orcid.org/0000-0003-3848-4242>

### Marta Cantero Gamito

Investigadora, Florence School of Transnational Governance, Instituto Universitario Europeo de Florencia; profesora de Derecho de las Tecnologías de la Información, Universidad de Tartu (Estonia). *marta.cantero@eui.eu*. ORCID: <https://orcid.org/0000-0001-8200-3456>

La inteligencia artificial (IA) y su desarrollo hiperacelerado suponen un desafío tecnológico, económico, social y geopolítico que los grandes poderes globales abordan desde diversas visiones estratégicas, a veces contradictorias. En este proceso de revolución tecnológica, la democracia y la IA se impactan mutuamente de muchas maneras. Esta última ocupa de manera creciente un lugar institucional en la administración pública y en las relaciones internacionales, desafiando la legitimidad e incluso la racionalidad del diseño de procesos para la toma de decisiones, lo que determina nuevos centros de poder. Así, la IA no solo está cambiando la relación democrática entre gobernantes y gobernados, sino también el reordenamiento entre poderes globales. Asistimos a un proceso acelerado de colonización algorítmica de la esfera pública y de la propia democracia, pero también de los recursos naturales imprescindibles para el desarrollo y funcionamiento de la IA.

En este contexto, el número 138 de *Revista CIDOB d'Afers Internacionals* pretende contribuir a la discusión pública sobre los retos que conlleva una IA democrática, tanto en su dimensión en línea (*online*) como fuera de ella (*offline*), porque la discusión sobre cómo deben articularse categorías, grupos e individuos para hacer frente a las discriminaciones que proceden de la pertenencia a un determinado colectivo, no depende únicamente de la agregación algorítmica, sino de la discusión política en general. De hecho, la IA y los sistemas de gobernanza algorítmica en general han acabado reproduciendo sesgos discriminatorios preexistentes. Se ha demostrado que la elaboración de perfiles impulsada por IA y la toma de decisiones está sesgada en contra de ciertos grupos étnicos y minoritarios, por lo que la discriminación de datos es ya un problema social real.

¿Cómo podemos lograr una IA capaz de adaptarse, aprender y evolucionar en respuesta a las discusiones políticas complejas y dinámicas que caracterizan la democracia? ¿Cómo se puede promover la gobernanza de una IA convertida en herramienta de poder global? ¿Cómo podemos transformar los paradigmas dominantes sobre el uso de la tecnología a fin de favorecer el desarrollo de una IA que no esté basada únicamente en el razonamiento lineal y deductivo? Con estas preguntas en mente, el monográfico que presentamos reflexiona sobre casos específicos de aplicación de la IA y la gobernanza algorítmica, así como sobre los riesgos que estas comportan.

## Riesgos para la integridad de la democracia

Daniel Innerarity explora en su artículo los límites de la gobernanza algorítmica y de las expectativas que suscita. Según sus palabras, «los algoritmos realizan una doble promesa de objetividad y subjetividad, es decir, de neutralidad ideológica y, al mismo tiempo, de respeto absoluto a nuestros deseos». Sin embargo, no se trata solo de eficacia, ni de un mero procesamiento de información, sino de su interpretación en un contexto de pluralismo garantizado. El gobierno democrático no consiste en proporcionar ciertos *outputs*, sino en garantizar que todos los ciudadanos puedan tomar parte en el proceso de formación de la voluntad política y en los procesos de decisión del gobierno.

La aportación de Anna Levitina también analiza la responsabilidad humana en este proceso de algoritmización, pero lo hace desde una perspectiva distinta, en el marco del Reglamento General de Protección de Datos (RGPD) y la Ley de la IA de la Unión Europea (UE). El primero, en vigor desde mayo de 2018, estableció un marco fundacional para la participación humana en la toma de decisiones automatizada (ADM, por sus siglas en inglés), exigiendo una implicación

humana significativa en estos procesos. Sin embargo, la autora considera que la legislación actual es insuficiente para abordar los aspectos cualitativos de la supervisión humana y de integrar a las personas supervisoras en los modelos de gobernanza.

Por su parte, Carlos Saura y Patrici Calvo examinan las principales estrategias vinculadas con las tecnologías digitales disruptivas para la manipulación de la opinión pública a través de la potenciación del desarrollo de un patrón totalmente ficticio, creado de forma artificial en favor de intereses determinados. El artículo aborda el impacto de este nuevo contexto democrático *algoritmizado, sintetificado* y masivo en una opinión pública y unos sistemas democráticos expuestos a esta suerte de despotismo tecnológico.

## Riesgos para la integridad de colectivos vulnerables

El proceso de *plataformización*, ya sea de las infraestructuras de datos, los mercados o la gobernanza, ha amplificado la vulnerabilidad de determinados colectivos. Al respecto, el artículo de Águeda Gómez Suárez y Rosa M<sup>a</sup> Verdugo Matés examina cómo el ciberespacio se ha transformado en un nuevo dispositivo de dominación a través de la difusión y proliferación de contenidos misóginos y sexistas en la red (ciberviolencia, «pornografía *mainstream*», industria de la explotación sexual, «proxenetismo digital»), que contribuyen a perpetuar la jerarquía sexual y la «monetización de la misoginia».

Asimismo, la introducción de los sistemas de IA en el control y la gestión migratorios europeos conlleva un proceso avanzado de *securitización* tecnológica de las fronteras, que ha creado nuevas vulnerabilidades y dinámicas discriminatorias. Sobre este tema, la contribución de Encarnación La Spina analiza por qué el nuevo marco regulador de la IA aprobado en la UE es insuficiente para evitar el impacto discriminatorio de sesgos, estereotipos y riesgos asociados a las dinámicas asimétricas de poder y a las estructuras de injusticia social subyacentes a las migraciones.

Por su parte, Marta Cantero Gamito explora cómo la IA está reconfigurando el acceso a la justicia, al analizar tanto las oportunidades como los riesgos asociados a la búsqueda de accesibilidad y la eficiencia. Al igual que otros servicios *low-cost*, este entendimiento de la justicia en términos de eficiencia presenta el riesgo de diluir la calidad y el rigor que tradicionalmente se espera del sistema judicial en un Estado de derecho.

En este contexto, Miguel Guillén Burguillos y Eloi Serrano Robles proponen integrar los principios de la economía social para promover un desarrollo de la IA que incorpore la ética, la justicia, la gestión democrática, la equidad, la sostenibilidad y la defensa de los intereses de las personas, sus comunidades y sociedades, más allá de los beneficios económicos.

Y en esta aproximación al impacto colectivo de la IA, tampoco las ciudades han escapado a la imparable *plataformización* que impone un nuevo régimen de gubernamentalidad a nivel local que opera a partir del uso masivo e intensivo de algoritmos de IA. José Luís Blasco Ejarque y Francisco Javier Tirado examinan la evolución de las denominadas «ciudades inteligentes» (*smart cities*) hacia ciudades progresivamente «posinteligentes», donde el poder algorítmico se vuelve opaco e inteligible en cuanto nos acercamos a sus mecanismos de funcionamiento internos.

## Riesgos geopolíticos para una gobernanza inclusiva

El debate sobre la gobernanza de la IA va de la mano de la carrera geopolítica por liderar la innovación tecnológica. Por este motivo el número 138 de *Revista CIDOB d'Afers Internacionals* también busca reflexionar sobre el impacto de la IA en la transformación de las relaciones internacionales y la seguridad. La irrupción y despliegue de la IA tiene un impacto exponencial en el proceso de transformación y reordenamiento geopolítico global, en la creación de nuevas dependencias y en el extractivismo de recursos naturales. En un contexto de confrontación, donde Estados Unidos, China y, en menor medida, la UE lideran la llamada «carrera» por el desarrollo de la IA, Carme Colomina Saló y Marta Galceran-Vercher analizan los espacios, procesos y actores, actualmente invisibilizados en los discursos dominantes sobre la geopolítica de la IA, y que son imprescindibles para entender cómo se van a desarrollar los sistemas algorítmicos en los próximos años.

No cabe duda de que la ubicuidad de la IA ya es indiscutible, pero su gobernanza aún está en construcción. En este contexto, este número aporta una mirada propia, compleja y heterogénea al debate sobre cómo y por qué es urgente democratizar la IA.

# Defensa y crítica de la gobernanza algorítmica

## Defence and criticism of algorithmic governance

### Daniel Innerarity

Catedrático de Filosofía Política, investigador Ikerbasque, Universidad del País Vasco; titular de la cátedra de Inteligencia Artificial y Democracia, Instituto Europeo de Florencia. [dinner@ikerbasque.org](mailto:dinner@ikerbasque.org). ORCID: <https://orcid.org/0000-0003-4307-8468>

**Cómo citar este artículo:** Innerarity, Daniel. «Defensa y crítica de la gobernanza algorítmica». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 11-25. DOI: doi.org/10.24241/rcai.2024.138.3.11

**Resumen:** El problema de la relación entre gobernanza algorítmica y democracia representa una continuidad y, al mismo tiempo, una ruptura con las clásicas formas de administración burocrática. Para entender su compatibilidad con los valores democráticos, hay que examinar el alcance de las expectativas que suscita, básicamente en cuanto a proporcionar más objetividad a las decisiones políticas y adoptarlas con una mayor consideración de nuestra subjetividad como ciudadanos destinatarios de tales decisiones. Los límites de estas promesas nos obligan a concluir en la inevitabilidad de la decisión humana, de la política en cualquier entorno tecnológico, incluido el configurado por las nuevas formas de gobernanza algorítmica.

**Palabras clave:** democracia, gobernanza algorítmica, inteligencia artificial (IA), decisión política, burocracia

**Abstract:** The problem of the relationship between algorithmic governance and democracy represents at once continuity and a break with traditional forms of bureaucratic administration. To understand its compatibility with democratic values, it is necessary to examine the scope of the expectations it raises, basically related to bringing more objectivity to political decisions and adopting them with greater consideration for our subjectivity as citizens who are the targets of these decisions. The limits of these promises force us to conclude on the inevitability of human decision-making, of politics in any technological environment, including the one shaped by the new forms of algorithmic governance.

**Key words:** democracy, algorithmic governance, artificial intelligence (AI), political decision-making, bureaucracy

## De la burocracia a la gobernanza algorítmica

En cuanto una comunidad política alcanza un cierto nivel de complejidad aparece la necesidad de objetivar y automatizar las decisiones colectivas. A partir del momento en el que el número de actores y factores que intervienen sobrepasa las capacidades individuales y centralizadas, las decisiones se vuelven más procedimentales y menos carismáticas. Muchos de los interrogantes que se plantean acerca de la justicia y democraticidad de las decisiones adoptadas mediante sistemas algorítmicos ya surgieron con ocasión de las decisiones burocráticas o mediante el recurso al saber experto. La burocracia nació en buena medida como respuesta a la incapacidad de gestionar la creciente complejidad del mundo en entornos desconocidos, pero también desde la promesa de superar la subjetividad, arbitrariedad e inconsistencia de las decisiones humanas.

Cuando se plantea una incompatibilidad entre decisiones estandarizadas del tipo que sea y consideraciones humanistas, no hay que olvidar que estos procedimientos se inventaron precisamente para minimizar la intervención humana en la toma de decisiones. Porter (1995) denominó «culto a la impersonalidad» a aquella cultura de la cuantificación en la que se aspira a reducir el elemento humano todo cuanto sea posible: principios formalizables frente a la interpretación subjetiva, estándares unitarios en lugar de caos metodológico y dominio del derecho en vez de poder humano. En este nuevo continente de la objetividad, reinarían una «objetividad mecánica» y una ciencia desinteresada que dejarían fuera todo lo que fuese personal, indiosincrático o perspectivista; ya no se confía en la integridad de los que dicen la verdad o en el prestigio de instituciones ejemplares, sino en procedimientos fuertemente estandarizados (ibídem). La fórmula más radical para expresarlo podría ser esta: «en vez de libertad de la voluntad, las máquinas ofrecerían liberarse de la voluntad» (Daston y Galison, 2010: 49). Esa esperanza hacia los datos y la objetividad aumenta en una cultura política y social caracterizada por la desconfianza, las crisis y la incertidumbre; el recurso a una cierta objetividad beneficia tanto a gobernantes como a gobernados, protege a quienes toman decisiones y genera confianza en quienes son afectados por ellas.

La era digital ha acentuado esta vieja tendencia. Gobernar es ya en gran medida –y lo será aún más– un acto algorítmico; una buena parte de las decisiones de gobierno son adoptadas por sistemas automatizados. Esta manera de gobernar ha sido definida de diversas maneras: «el poder está cada vez más en el algoritmo» (Lash, 2007: 71); «la autoridad se expresa cada vez más mediante algoritmos» (Pasquale, 2015: 1). Este sistema en el que se utilizan algoritmos para recoger, cotejar y organizar los datos a partir de los cuales se toman las decisiones ha sido denominado «algocracy» (Aneesh, 2009; Danaher, 2016), «algorith-

mic governmentality» (Rouvroy, 2013), «algorithmic management» (Lee *et al.*, 2015), «algorithmic regulation» (Yeung, 2017) o «governance by algorithms» (Just y Latzer, 2017). La era digital añade, ahora, una nueva promesa a la acción político-administrativa: mientras que las burocracias estatales se basan en las estadísticas y la información cuantificada, las nuevas técnicas analíticas ofrecen una mejora de los métodos anteriores; si los análisis de datos tenían un alto coste de tiempo y dinero, los actuales son rápidos y baratos; donde antes había muestras de la sociedad, hoy se dispone de datos de los grupos sociales enteros; anteriormente se necesitaban teorías y, en la actualidad, las cantidades de datos hablan por sí mismas; donde se medía con criterios humanos, ahora disponemos de la objetividad de algoritmos agnósticos.

El recurso a algoritmos y decisiones automatizadas responde a la necesidad de hacer frente a diversas formas de complejidad, como la identificación de las distintas perspectivas e intereses de una sociedad cada vez más plural, o la eficiente provisión de servicios públicos. La gobernanza algorítmica potencia enormemente las capacidades de gestión a través de grandes cantidades de datos y en relación con problemas complejos. De este modo, no solo el mundo parece habérsenos hecho más legible, sino que se han abierto nuevas posibilidades de intervención política, una mayor eficiencia, una más inteligente regulación y una más temprana anticipación de determinados problemas. Se promete así una acción de gobierno que reduciría la complejidad de los fenómenos sociales a una medida aceptable.

El incremento de sistemas de decisión conducidos por algoritmos y datos significa que las máquinas no solo apoyan a los humanos en sus decisiones, sino que, en algunos casos, incluso los sustituyen, ya sea en parte o completamente. La cuestión que todo esto plantea es hasta qué punto y de qué modo la utilización de sistemas de decisión automatizada (ADS, por sus siglas en inglés) es compatible con lo que consideramos un sistema *político* de toma de decisiones. ¿Qué significa realmente la introducción masiva de procedimientos de decisión automatizada para la acción de gobierno? ¿Es acorde este tipo de gobernanza con la democracia? De la democracia se espera que responda a la expectativa de ser un verdadero autogobierno del pueblo y, al mismo tiempo, que el sistema político resuelva eficazmente los problemas de la sociedad. La integración de este doble objetivo de la sociedad –democracia y eficacia– no parece algo evidente, sino en tensión. ¿Hasta qué punto cumple esta doble promesa la gobernanza algorítmica? ¿Cómo compatibilizar la heterogeneidad de preferencias, valores e intereses con la operatividad del sistema político? Para responder a esta cuestión es necesario distinguir las diversas funciones o momentos de la política y examinar la aptitud de estos procedimientos para realizar esas tareas sin dañar los principios democráticos.

## Las expectativas democráticas de la gobernanza algorítmica

Los algoritmos realizan una doble promesa de objetividad y subjetividad, es decir, de neutralidad ideológica y, al mismo tiempo, de respeto absoluto a nuestros deseos. Se trata de dos promesas que tienen unos efectos muy beneficiosos sobre la política democrática, pues permiten una valoración más objetiva de las políticas públicas y un mejor conocimiento de las preferencias sociales; sin embargo, también tienen sus límites e inconvenientes.

### La promesa de objetividad

Resulta muy seductora la promesa de la decisión algorítmica: no se trata solo de ahorrar tiempo y dinero, sino de promover la objetividad. Ya en 1976, Joseph Weizenbaum (1976: 108), uno de los pioneros de la investigación en inteligencia artificial (IA), defendía este valor

**Los algoritmos realizan una doble promesa de objetividad y subjetividad, es decir, de neutralidad ideológica y, al mismo tiempo, de respeto absoluto a nuestros deseos. Se trata de dos promesas que tienen unos efectos muy beneficiosos sobre la política democrática, pues permiten una valoración más objetiva de las políticas públicas y un mejor conocimiento de las preferencias sociales; sin embargo, también tienen sus límites e inconvenientes.**

asegurando que una computadora no podía ser seducida por la mera elocuencia. Al minimizar la presencia humana, los algoritmos hacen que las decisiones sean menos vulnerables a nuestros sesgos y sentimientos (Sandvig, 2015; Zarsky, 2016). Los algoritmos suelen percibirse como objetivos y sus evaluaciones como justas, precisas y libres de subjetividad, errores y pretensiones de poder; es más, su «objetividad» es lo que les proporciona legitimación como mediadores de conocimiento relevante; no son solo instrumentos para decidir, sino también estabilizadores de la confianza; aseguran que «las valoraciones son precisas y justas, sin fallos, subjetividad o distorsiones» (Gillespie, 2014: 79; Mager, 2012).

Si, como afirmaba Lindblom (1965), la esencia de la cultura democrática es el incrementalismo y la comparación, el ensayo y error, los gobiernos cuentan ahora con instrumentos sofisticados para medir la efectividad de sus políticas públicas, asegurar su implementación y valorar sus resultados. En vez de la planificación centralizada, dominada por expertos y burócratas,

la tecnología nos permite introducir criterios de valoración más dispersos y competitivos, menos ideologizados. La puesta en marcha de ADS se justifica porque con su ayuda las decisiones no son solo más eficientes, sino también menos partidistas y más justas. Se abre paso así la idea de que los sistemas que deciden sin influencias humanas pueden ser más neutros y objetivos (Martini y Nink, 2017). Así, los científicos de datos remplazarían a los expertos (Chen *et al.*, 2014: 205). Tendríamos unos instrumentos que parecen satisfacer la esperanza de proporcionar una mayor racionalidad a los procesos de decisión y contrarrestar la subjetividad y los prejuicios ideológicos o de cualquier tipo que suelen motivar muchas de las decisiones humanas. «La calculadora ideal es un ordenador, ampliamente venerado en parte porque es incapaz de subjetividad» (Porter, 1995: 47). De este modo, seríamos capaces de dejar atrás los pronunciamientos ideológicos sin transformaciones efectivas de la realidad social.

Esta pretensión no es del todo nueva, al igual que tampoco lo es su crítica. La idea de autoridad burocrática de Max Weber ya había ensalzado los valores de eficiencia y objetividad, aunque también había advertido de sus límites, así como de que otro tipo de autoridades podían surgir precisamente en virtud del ideal de objetividad. En principio, todas las tendencias patológicas de las clásicas burocracias se aplican también a las decisiones automatizadas (Peeters y Widlak, 2018). Baste con advertir que la gente tiende a aceptar con demasiada facilidad que los procesos automatizados son verdaderos y precisos (Citron, 2007). El hecho de que una decisión sea el resultado de un proceso automatizado parece conferirle una legitimidad que sería el resultado de su neutralidad, debilitando así la exigencia de justificación (Gillespie, 2016: 27). De este modo, podríamos caer en la «falacia de la neutralidad», consistente en pensar que el *machine learning* proveya un tratamiento más igualitario y objetivo de los individuos (Sandvig, 2015). Desde que se formularon las pretensiones de objetividad –en el entorno burocrático y en la era digital–, no ha dejado de advertirse que tales procedimientos no cumplen con esa promesa, ya que que generan otro tipo de distorsiones y no están exentas de arbitrariedad; los algoritmos, a menudo, reflejan e incluso potencian los prejuicios que están profundamente asentados en una sociedad.

**Desde que se formularon las pretensiones de objetividad –en el entorno burocrático y en la era digital–, no ha dejado de advertirse que tales procedimientos no cumplen con esa promesa, ya que que generan otro tipo de distorsiones y no están exentas de arbitrariedad; los algoritmos, a menudo, reflejan e incluso potencian los prejuicios que están profundamente asentados en una sociedad.**

## La promesa de subjetividad

El segundo vector de democratización vendría del conocimiento de la voluntad real de la gente a la que un Gobierno democrático debe servir; reforzándose la cadena de legitimación en la medida en que permitiría tomar como punto de partida las decisiones reales de las personas, únicamente a partir de las cuales se puede configurar la voluntad popular. En un mundo lleno de sensores, algoritmos, datos y objetos inteligentes se configura una suerte de *sensorium* social que permite personalizar los servicios de salud, transporte y energía. Gracias a la ingeniería de los datos, nos estamos moviendo hacia una comprensión cada vez más granular de las interacciones individuales y unos sistemas más capaces de responder a las necesidades individuales. En virtud de la microsegmentación y granularidad, podemos disponer de una sociedad «algorithmically attuned», de manera que los deseos que la ciudadanía expresa de hecho en su comportamiento cotidiano pueden ser conocidos con un altísimo grado de exactitud. De esta forma, a la objetividad de los métodos de gobernanza algorítmica le correspondería una mayor subjetividad en sus destinatarios, quienes verían mejor reconocida, respetada y satisfecha su particularidad.

Los sistemas algorítmicos sirven para categorizar a los individuos y prever sus preferencias a partir de una gran cantidad de datos acerca de ellos. El modelo de negocio de muchas empresas digitales se apoya en el hecho de que conocen a los usuarios mejor incluso que ellos mismos y, en virtud de la previsión de su comportamiento, les ofrecen lo más adecuado en el momento oportuno. De este modo, recibiríamos lo que supuestamente queremos, algo que corrige el hecho de que tantas veces no sepamos propiamente qué es aquello que queremos. Y lo que así se ofrece, no parece que limite o contravenga nuestra autodeterminación. El cómodo paternalismo de las sociedades algorítmicas consiste en que da a las personas lo que estas quieren, gobierna con incentivos proporcionados y se adelanta, invita y sugiere. Trasladar este modelo a la política no tendría mayores problemas si no fuera porque el precio de estas prestaciones suele ser el sacrificio de alguna esfera de libertad personal. Teniendo en cuenta que hay una discrepancia en la autodeterminación que supuestamente exigimos y la que de hecho estamos dispuestos a ejercer cuando hay comodidades y prestaciones de por medio, el resultado es que la satisfacción de necesidades se hace con frecuencia a cambio de espacios de libertad. Estudios empíricos demuestran que los humanos infravaloramos los peligros que los ADS representan para nuestra libertad y derechos personales (Wouters *et al.*, 2019). Aunque es cierto que así se satisfacen muchos de nuestros deseos, ello es a cambio de una cierta renuncia a reflexionar sobre ellos; lo que queremos se sitúa por encima de lo que queremos

querer, y la voluntad mínima e implícita del consumidor sustituye a la voluntad política explícita. Detrás de la gobernanza algorítmica hay una concepción de la vida social como si en ella no hubiera fallos ni crisis, de manera que tampoco pudiera ponerse a prueba sus prestaciones ni realizarse cuestionamientos de las normas establecidas.

## **Las limitaciones democráticas de la gobernanza algorítmica**

La idea de Alain Turing del ordenador como una «máquina universal» no significa que esta valga para resolver cualquier problema; es un error creer que las tecnologías digitales pueden encargarse de todos los problemas políticos y sociales. La gobernanza algorítmica es muy adecuada para mejorar ciertos aspectos del proceso político, pero resulta de escasa utilidad para otros; puede corregir deficiencias y sesgos humanos, sirve para identificar determinadas preferencias, para medir los impactos, pero es inadecuada para aquellas dimensiones del proceso político que no son susceptibles de computación y optimización, áreas que no tienen una fácil cuantificación y medida, o sea, para el momento genuinamente democrático en el que se deciden los criterios y objetivos que posteriormente la tecnología puede optimizar. De acuerdo con la taxonomía elaborada por Misuraca y Van Noordt (2020), la IA podría considerarse muy útil para seis tipos de desafíos de gobierno: asignación de recursos, análisis de grandes conjuntos de datos, superación de la carencia de expertos, predicción de escenarios, gestión de tareas procedimentales y repetitivas, así como agregación y resumen de datos diversos (Duberry, 2022).

La razón de que los algoritmos sean políticamente limitados reside en su carácter instrumental. Los algoritmos sirven para conseguir objetivos predeterminados, pero ayudan poco a determinar esos objetivos, tarea propia de la voluntad política, de la reflexión y deliberación democrática. La función de la política es decidir el diseño de las estrategias de optimización algorítmica y mantener siempre la posibilidad de alterarlas, especialmente en entornos cambiantes. En una democracia todo debe estar abierto a momentos de repolitización, es decir, a la posibilidad de cuestionar los objetivos establecidos, las prioridades y los medios. Para esto es para lo que sirve la política y para lo que no sirven los algoritmos. El gobierno algorítmicamente optimizado no tiene capacidad para resolver los conflictos propiamente políticos o la dimensión política de esos conflictos, es decir, cuando están en cuestión los marcos, fines o valores. Como decía Lucy Suchman

(2007) en otro contexto, los robots actúan muy bien cuando el mundo ha sido dispuesto (*arranged*) del modo en que debía ser dispuesto.

Puede ilustrar esta dualidad de fines y medios, de objetivos políticos y estrategias de optimización algorítmica, el sistema de distribución de los alumnos que se puso en marcha para las escuelas de la ciudad de Nueva York y el debate correspondiente acerca de qué valores priorizar en esa distribución (Krüger y Lischka, 2018). El sistema puede priorizar la satisfacción al máximo de las preferencias individuales o una mezcla social equilibrada en las escuelas. Ambos objetivos cuentan con buenas razones a su favor; una opción favorece los deseos individuales y la otra la cohesión social. Pero también es discutible, si se quiere respetar al mismo tiempo los dos valores, qué grado de compromiso o equilibrio entre ellos parece más deseable y realizable. Para decidirlo, hace falta un debate político acerca de valores e implicar principalmente a los afectados, un debate del que no puede exonerarnos un algoritmo. En este y otros casos, no se trata solo de la implementación o la transparencia de los algoritmos utilizados, sino de juicios de valor sobre las posibilidades alternativas y definir los objetivos de la educación, que son diversos y en parte concurrentes, como corresponde a una

**La gobernanza algorítmica es muy adecuada para mejorar ciertos aspectos del proceso político, pero resulta de escasa utilidad para otros; puede corregir deficiencias y sesgos humanos, identificar determinadas preferencias, medir los impactos, pero es inadecuada para aquellas dimensiones del proceso político que no son susceptibles de computación y optimización, áreas que no tienen una fácil cuantificación y medida, o sea, para el momento genuinamente democrático en el que se deciden los criterios y objetivos que posteriormente la tecnología puede optimizar.**

saber consideramos relevante. Podría incluso afirmarse que, si la optimización como principio es algo deseable, la ideología de la optimización (pensar que la implementación eficaz de ciertos objetivos puede hacer innecesaria la discusión política acerca de tales objetivos) puede ser una estrategia de despolitización. Así entendida, la optimización es exactamente lo contrario de la política, que es más bien imaginación, anticipación, trascender el actual estado de hechos (Rouvroy, 2020).

sociedad pluralista. Los procesos de negociación política tienen prioridad sobre las soluciones técnicas, y estas no pueden sustituir a aquellos. Estamos, por tanto, ante ese tipo de asuntos que denominamos cuestiones políticas.

En sentido estricto, cuestiones políticas son aquellas que solo se pueden resolver con juicios de valor; las otras son cuestiones técnicas en las que se decide la implementación técnica de los objetivos pretendidos y sobre la base del saber disponible. En ocasiones, también es algo políticamente controvertido qué clase de optimización es la satisfactoria y qué

La gobernanza algorítmica se orienta a realizar objetivos que no han sido discutidos y que ella misma no establece ni pone en cuestión. Ahora bien, la política democrática no es un mero procesamiento de información, sino su interpretación en un contexto de pluralismo garantizado; no se trata de cómo realizar mejor ciertos objetivos, sino de cómo decidirlos. La resolución de problemas de carácter administrativo es muy distinta de la política entendida como el conflicto de interpretaciones acerca de la realidad, donde no se trata tanto de optimizar resultados como de establecerlos. A este respecto, podemos advertir que hay una gran diferencia entre cómo aprenden los sistemas algorítmicos y cómo se toman las decisiones democráticas (Hildebrandt, 2016). Los ADS procesan información para realizar lo mejor posible ciertos objetivos, mientras que la política democrática, en contraste, no trata de optimizar objetivos predefinidos sino sobre todo de averiguar cuáles deberían ser esos objetivos. El aprendizaje de los sistemas algorítmicos no puede remplazar al tipo de aprendizaje que tiene lugar en la política democrática. Lo político empieza allí donde se ha de debatir acerca de qué deben satisfacer los algoritmos, qué valores deben cumplir, a qué concepción de lo justo han de servir. Los sistemas algorítmicos de aprendizaje se diferencian de los procesos políticos democráticos en que estos últimos se caracterizan por una discusión abierta y continua que lleva a renegociar una y otra vez los objetivos y las decisiones, mientras que la gobernanza algorítmica los presupone y trata de alcanzarlos de una manera optimal y agregativa, sin cuestionarlos. Podría formularse esta idea recordando aquella afirmación de John von Neumann (1966: 51): podemos construir un instrumento capaz de hacer todo lo que puede ser hecho, pero no se puede construir un instrumento que nos diga si algo es factible. Dicho de otra manera: la decisión acerca de qué es computerizable no se puede a su vez computerizar.

Es lógico que una tecnología tan poderosa como la IA haya despertado expectativas similares a las que formularon los viejos positivimos, la tecnocracia o el proclamado final de las ideologías. Todos ellos declinaron de diversas maneras aquella seducción de «un mundo administrado» (Adorno y Horkheimer, 1988), y que ahora puede acogerse bajo el denominador de «la ideología cibernética», impulsada por la creencia de que es posible gobernar la realidad con plena eficacia, exactitud y previsibilidad (Nunes, 2011: 3). Esa manera de concebir la tecnología nos plantea el desafío de pensar hasta qué punto es posible entenderla como políticamente configurable. Antes de debatir acerca de la deseable regulación digital, deberíamos preguntarnos si es posible *politizar* este nuevo entorno, lo cual nos obliga, frente al neutralismo tecnológico, a identificar y hacer explícitas las valoraciones implícitas que subyacen en cualquier procedimiento de decisión que pretende una objetividad incontestable.

La ideología tecnocrática ha sido formulada de manera enfática, pero también es operativa de un modo más bien banal y de apariencia inofensiva. Me

refiero a ese lugar común según el cual no debería importarnos quién gobierna, si es de derechas o de izquierdas, sino que gestione bien, como si esa gestión pudiera valorarse sin recurrir a estimaciones ideológicas. Este tópico resulta plausible solo en la medida en que, efectivamente, derecha e izquierda ya no son lo que eran y como categorías rígidas cada vez explican menos. Pero muchas veces quien lo sostiene no suele estar deseando una política desideologizada sino una política despolitizada.

Como ocurre en la política en general, también cuando hablamos de gobernanza algorítmica la idea de producir mejores decisiones con la ayuda de máquinas requiere que haya previamente un criterio acerca de qué es una buena decisión (König y Wenzelburger, 2021). Los artefactos que se encargan de optimizar las decisiones no hacen innecesaria la discusión acerca de qué es

**El gobierno algorítmicamente optimizado no tiene capacidad para resolver los conflictos propiamente políticos o la dimensión política de esos conflictos, es decir, cuando están en cuestión los marcos, fines o valores.**

una buena decisión. Es cierto que la IA sirve para informar decisiones y optimizar resultados, pero, aunque algunos economistas hayan intentado cuantificar y medir el bienestar agregado, no hay una noción predefinida e incontestable de qué es un resultado satisfactorio en po-

lítica. La democracia es un sistema político que parte de la ignorancia acerca de qué pueda ser una buena decisión, que recela de que alguien pretenda saberlo y pone en marcha procedimientos de aprendizaje colectivo para superar esa perplejidad.

Es evidente que el análisis de datos y la gobernanza algorítmica proporcionan a los actores políticos un conocimiento y una capacidad de intervención muy valiosa. Lo que permanece abierto es el modo concreto en el que se lleva a cabo y al servicio de qué valores se pone. Buena parte de la función de la técnica es además multiplicar las alternativas posibles y ponerlas a nuestra disposición. La política está precisamente para decidir cuál de esas alternativas resulta más adecuada a la luz de lo que resulta socialmente preferible. También cuando se trata de algoritmos hay alternativas; no son soluciones funcionales indiscutibles. La gran promesa de la gobernanza algorítmica es que unos resultados óptimos nos hagan olvidar los procedimientos deseables. Es un tipo de gobernanza que parece preferir la efectividad, aunque sea al precio de ser excluidos de la toma de decisiones o reducidos a una presencia mínima, implícita e individual, bajo la forma de requerimientos y preferencias presentes en nuestras huellas digitales. Pero si la ciudadanía no puede supervisar ni controlar de algún modo las decisiones algorítmicas, no podemos llamar a eso autogobierno del pueblo.

Siendo muy importante los resultados del gobierno, lo definitorio de la democracia es más el procedimiento que el resultado. El gobierno democrático no consiste en proporcionar ciertos *outputs*, sino en garantizar determinados *inputs*, concretamente aquellos que aseguran la igual libertad de todos los ciudadanos para tomar parte en el proceso de formación de la voluntad política y en los procesos de decisión (Urbinati, 2014). Aunque la gobernanza algorítmica sea muy *responsive*, presupone una legitimidad *input* que ella misma no puede proporcionar (König, 2018: 289). A este tipo de gobernanza le falta una autorización colectiva, aunque disponga de una gran cantidad de información sobre las preferencias individuales. La gobernanza algorítmica únicamente puede ser democrática cuando sus objetivos y procedimientos han sido expresamente autorizados por el pueblo en un acto de naturaleza política

En cualquier caso, el trabajo analítico de distinguir los diferentes momentos del proceso de decisión político no debería llevarnos a pensarlos como completamente separados y sin ninguna relación entre sí. Del mismo modo que el momento deliberativo favorece la calidad de los resultados, la medición de los impactos reales permite valorar también la calidad de los procedimientos deliberativos. Que fines y medios sean diferentes no significa que no se condicionen mutuamente de alguna manera. El respeto de la complejidad política nos obliga a pensar de qué modo interactúan los momentos que, precisamente a causa de ese respeto, habíamos tenido que distinguir.

## Conclusión: la inevitabilidad de decidir

El gran desafío que nos plantea la era digital es el de resistir a los encantos de la despolitización de nuestras sociedades y superar la inercia de los modos de gobierno tradicionales, no dejarse seducir por el discurso falsamente apolítico o posideológico, pero, al mismo tiempo, evitar insistir en unas prácticas que no se corresponden en absoluto con las nuevas realidades sociales. Estamos ante un intento de concebir la sociedad de un modo despolitizado. Tim O'Reilly (2011), uno de los oráculos de Silicon Valley, inventor de los conceptos de «web 2.0» y «open source», plantea pensar el gobierno como una plataforma, o sea, extender el modelo de las aplicaciones comerciales a la administración de las cosas comunes. En nombre de una lucha contra los déficits democráticos y el exceso de burocracia, propone reducir el papel del Estado al de suministrador de acceso y plataforma, sobre la cual la ciudadanía podría definir por sí misma y con toda libertad sus prioridades políticas.

Si en un principio han sido los poderes públicos los que han impulsado el desarrollo tecnológico, ahora el movimiento sería el inverso: se invita a que el Estado se inspire en las plataformas, esto es, a no servir más que de infraestructura supuestamente neutra para las transacciones entre los individuos. En cualquier caso, entre la seducción de un mundo despolitizado y la inercia a mantener nuestras instituciones con la vieja cultura política, hay un amplio espacio para pensar el lugar que debe ocupar la política en estas nuevas realidades.

En este punto, el lugar que ocupe la decisión humana es crucial. He estudiado las razones de la democracia epistémica (Innerarity, 2020) y, estando de acuerdo en que las sociedades contemporáneas necesitan una enorme movilización cognitiva para hacer frente a los problemas que deben resolver,

**Hay que hacer todo lo posible para que las sociedades tomen las mejores decisiones, pero la legitimidad final no procede de la corrección de sus decisiones, sino del poder de decisión que tiene la ciudadanía con independencia del buen o mal uso que haga de este poder.**

mi conclusión es que el argumento último a favor de la democracia no es epistémico sino decisonal. Hay que hacer todo lo posible para que las sociedades tomen las mejores decisiones, pero la legitimidad final no procede de la corrección de sus decisiones, sino del poder de decisión que tiene la ciudadanía

con independencia del buen o mal uso que haga de este poder. Una democracia produce mejores decisiones que sus modelos alternativos, pero no debe su legitimidad última a la bondad de sus decisiones, sino a la autorización popular que está detrás de esas decisiones. La inevitabilidad de decidir es la justificación definitiva de que la democracia sea una forma de gobierno donde los legos tienen la última palabra sobre los expertos. No parece que haya hoy por hoy un dispositivo tecnológico que nos libere completamente de esta necesidad de decidir.

Los procedimientos de la IA no pueden exonerarnos de esa decisión. Hay política allí donde, pese a toda la sofisticación de los cálculos, nos vemos finalmente obligados a tomar una decisión que no está precedida por razones abrumadoras ni conducida por unas tecnologías infalibles. Todos los procesos de tecnificación tienden a modelizar o automatizar de manera que el «factor humano» sea menos relevante. Los humanos no hemos dejado de soñar en «la perfecta tecnología de la justicia» (Lessing, 1999), pero tampoco hemos dejado de experimentar el peso de que sean nuestras decisiones las que carguen con la última responsabilidad de hacer que la sociedad sea justa. Un mundo humano tiene que ser un mundo negociable.

## Referencias bibliográficas

- Adorno, Theodor W. y Horkheimer, Max. *Dialektik der Aufklärung*, Berlin: Fischer, 1988.
- Aneesh, Aneesh. «Global Labor: Algorocratic Modes of Organization». *Sociological Theory*, vol. 27, n.º 4 (2009), p. 347-370.
- Chen, Min; Mao, Shiwen y Liu, Yunhao. «Big Data. A Survey». *Mobile Networks and Applications*, vol. 19, (2014), p. 171-209.
- Citron, Danielle Keats. «Technological Due Process». *Washington University Law Review*, vol. 85, n.º 6 (2007), p. 1.249-1.313.
- Danaher, John. «The threat of algocracy: Reality, resistance and accommodation». *Philosophy & Technology*, vol. 29, (2016), 245-268.
- Daston, Lorraine y Galison, Peter. *Objectivity*. Nueva Jersey: Princeton University Press, 2010.
- Duberry, Jérôme. *Artificial Intelligence and Democracy. Risks and Promises of AI-Mediated Citizen–Government*. Cheltenham: Elgar, 2022.
- Gillespie, Tarleton. «The Relevance of Algorithms», en: Gillespie, Tarleton; Boczkowski, Pablo J. y Foot, Kirsten A. (eds.) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA.: The MIT Press, 2014, p. 167-193.
- Gillespie, Tarleton. «Algorithm», en: Peters, Benjamin (ed.) *Digital Keywords: A Vocabulary of Information Society and Culture*. Nueva Jersey: Princeton University Press, 2016, p. 18-30.
- Hildebrandt, Mireille. «Law as Information in the Era of Data-Driven Agency: Law as Information». *The Modern Law Review*, vol. 79 , n.º 1 (2016), p. 1-30.
- Innerarity, Daniel. *Una teoría de la democracia compleja. Gobernar en el siglo XXI*, Galaxia-Gutenberg: Barcelona, 2020.
- Just, Natascha y Latzer, Michael. «Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet». *Media, Culture & Society*, vol. 39, n.º 2 (2017), p. 238-258.
- König, Pascal. «Algorithmen und die Verwaltung sozialer Komplexität. Zur Neukonfigurierung der Idee der Selbstregierung des Volkes». *Zeitschrift für Politikwissenschaft*, vol. 28, (2018), p. 289-312.
- König, Pascal y Wenzelburger, Georg. «Between technochauvinism and human-centrism: Can algorithms improve decision-making in democratic politics?». *European Political Science*, vol. 21, (2021), p. 132-149.
- Krüger, Julia y Lischka, Konrad. «Was zu tun ist, damit Maschinen den Menschen dienen», en: Mohabbat Kar, Resa; Thapa, Basanta y Parycek, Peter (eds.)

- (Un)berechenbar? *Algorithmen und Automatisierung in Staat und Gesellschaft*. Berlin: Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS, 2018, p. 440-470.
- Lash, Scott. «Power after hegemony». *Theory, Culture & Society*, vol. 24, n.º 3 (2007), p. 55-78.
- Lee, Min Kyung; Kusbit, Daniel; Metsky, Evan y Dabbish, Laura. «Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers», en: Begole, Bo y Kim, Jinwoo (eds.) *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Nueva York: ACM, 2015, p. 1.603-1.612.
- Lessing, Lawrence. *Code and other Laws of Cyberspace*. Nueva York: Basic Books, 1999.
- Lindblom, Charles. *The Intelligence of Democracy: Decision Making Through Mutual Adjustment*. Nueva York: The Free Press, 1965.
- Mager, Astrid. «Algorithmic Ideology: How Capitalist Society Shapes Search Engines». *Information, Communication & Society*, vol. 15, n.º 5 (2012), p. 769-787.
- Martini, Mario y Nink, David. «Wenn Maschinen entscheiden». *Neue Zeitschrift für Verwaltungsrecht*, n.º 10 (2017), 1-14.
- Misuraca, Gianluca y Van Noordt, Colin. «AI Watch – artificial intelligence in public services: Overview of the use and impact of AI in public services in the EU». *Publications Office of the European Union*, JRC Working Papers (JRC120399) (2 de julio de 2020) (en línea) <https://publications.jrc.ec.europa.eu/repository/handle/JRC120399>
- Neumann, John von. *Theory of Self-Reproducing Automata*. Urbana: University of Illinois Press, 1966.
- Nunes, Mark. *Error: Glitch, Noise, and Jam in New Media Culture*. Nueva York: Continuum, 2011.
- O'Reilly, Tim. «Government as a Platform». *Innovations: Technology, Governance, Globalization*, vol. 6, n.º 1 (2011), p. 13-40.
- Pasquale, Frank. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press, 2015.
- Peeters, Rik y Widlak, Arjan. «The digital cage: Administrative exclusion through information architecture – The case of the Dutch civil registry's master data management». *Government Information Quarterly*, vol. 35, n.º 2 (2018), 175-183.
- Porter, Theodore. *Trust in numbers: The pursuit of objectivity in science and public life*. Nueva Jersey: Princeton University Press, 1995.
- Rouvroy, Antoinette. «The end(s) of critique: data-behaviourism vs. due process», en: Hildebrandt, Mireille y de Vries, Katja (eds.) *Privacy, Due Process*

- and the Computational Turn. Philosophers of Law Meet Philosophers of Technology*. Nueva York: Routledge, 2013, p. 143-167.
- Rouvroy, Antoinette. «Algorithmic Governmentality and the Death of Politics». *Society, Media and Culture*, (20 de marzo de 2020) (en línea) <https://www.greeneuropeanjournal.eu/algorithmic-governmentality-and-the-death-of-politics/>
- Sandvig, Christian. «Seeing the sort: The aesthetic and industrial defense of ‘the algorithm’». *Journal of the New Media Caucus*, vol. 10, n.º 3 (2015), p. 1-21.
- Suchman, Lucy. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press, 2007.
- Urbinati, Nadia. *Democracy disfigured: opinion, truth, and the people*. Cambridge, MA.: Harvard University Press, 2014.
- Weizenbaum, Joseph. *Computer Power and Human Reason*. San Francisco: Freeman, 1976.
- Wouters, Niels; Kelly, Ryan; Velloso, Eduardo; Wolf, Katrin; Shahid Ferdous, Hasan; Newn, Joshua; Joukhadar, Zaher y Vetere, Frank. «Biometric Mirror: Exploring Values and Attitudes towards Facial Analysis and Automated Decision-Making». *Association for Computing Machinery*, Proceedings of the 2019 Designing Interactive Systems Conference, Nueva York, (23-28 de junio de 2019), p. 447-461.
- Yeung, Karen. «Algorithmic Regulation: A Critical Interrogation». *Regulation & Governance*, vol. 12, n.º 6 (2017), p. 1-19.
- Zarsky, Tal. «The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making». *Science, Technology and Human Values*, vol. 41, n.º 1 (2016), p. 118-132.



DOSSIER

**Presentación del dossier**

**El derecho al aborto: un tema pendiente en la región**

María Rosa Cevallos-Castells, Soledad Varea-Viteri y Sonia Ariza-Navarrete

**Disputas por la gobernanza de la autonomía reproductiva en América Latina y el Caribe**

Martina Ferretto

**“Yo también soy feminista”. Neoconservadurismo de género desde el activismo antiaborto en Chile**

María Angélica Cruz, Félix Aguirre-Díaz y Paula Eguren-Álvarez

**“Poner el cuerpo”: primeras experiencias de abortos ambulatorios con misoprostol en Argentina**

Natacha Mateo

**Lideresas indígenas y acceso al aborto en Ecuador: una mirada desde el ciberfeminismo y la ética del cuidado**

María Rosa Cevallos-Castells y Soledad Varea-Viteri

**Aborto por violación en redes sociales de Brasil: entre narrativas exitosas y dicotomías morales**

Mariana Prandini Assis y Nara Menezes Santos

TEMAS

**Familia, género y violencia económica: incumplimiento de las pensiones de alimentos en Chile**

Martina Yopo-Díaz y Javiera Fuentes-Landaeta

**Nuevos residentes, nuevos paladares: transformaciones espaciales y comerciales en Ciudad de México**

Raul Anthony Olmedo-Neri

**Organización del trabajo y protección social en cooperativas argentinas**

Maricel Massera

Número anterior:

**ICONOS 79: Alimentación y gentrificación: aproximaciones desde Latinoamérica**

Número siguiente:

**ICONOS 81: Transiciones juveniles al mercado laboral en América Latina**

Íconos. Revista de Ciencias Sociales está incluida en los siguientes índices científicos: Academic Search Premier; Directory of Publishing Opportunities (CABELL'S); Clasificación Integrada de Revistas Científicas (CIRC); Citas Latinoamericanas en Ciencias Sociales (CLASE); DIALNET; Directory of Open Access Journal (DOAJ); Emerging Source Citation Index (ESCI) Web of Science; ERIHPLUS; FLACSO Andes; Fuente Académica Plus; Hispanic American Periodical Index (HAPI); International Bibliography of the Social Science (IBSS); Informe Académico Thompson Gale; LatAm-Studies; LATINDEX catálogo; MIAR; Political Science Complete; REDALYC; REDIB; SciELO Ecuador; Sociological Abstracts; SCOPUS; Social Science Journals. Sociology Collection; Ulrich's Periodical Directory; Worldwide Political Science Abstracts (WPSA).

Página web: [www.revistaiconos.ec](http://www.revistaiconos.ec)

Correo electrónico: [revistaiconos@flacso.edu.ec](mailto:revistaiconos@flacso.edu.ec)

TEMAS

**La reproducción del estigma territorial.**

**Dinámicas socioespaciales de una colonia popular en México**

Miguel Ángel Monteverde-Ávalos

**Violencia sexual como violencia política en las dictaduras de Argentina y Chile**

Nadia Poblete-Hernández



FLACSO  
ECUADOR

Revista de la Facultad Latinoamericana de Ciencias Sociales - Sede Ecuador

Información y colaboraciones: ([revistaiconos@flacso.edu.ec](mailto:revistaiconos@flacso.edu.ec))

Sitio web: [www.revistaiconos.ec](http://www.revistaiconos.ec)

# Las otras geopolíticas de la inteligencia artificial

## The other geopolitics of AI

### Carme Colomina Saló

Investigadora sénior y editora, CIDOB; profesora asociada, College of Europe (Bélgica).  
ccolomina@cidob.org. ORCID: <https://orcid.org/0000-0003-3848-4242>

### Marta Galceran-Vercher

Investigadora sénior, CIDOB; profesora asociada, Universitat Pompeu Fabra (Barcelona).  
mgalceran@cidob.org. ORCID: <https://orcid.org/0000-0003-4991-1369>

**Cómo citar este artículo:** Colomina Saló, Carme y Galceran-Vercher, Marta. «Las otras geopolíticas de la inteligencia artificial». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 27-50. DOI: doi.org/10.24241/rcai.2024.138.3.27

**Resumen:** La irrupción y despliegue de la inteligencia artificial (IA) tiene un impacto exponencial en el proceso de transformación y reordenamiento geopolítico global. Si bien Estados Unidos, China y, en menor medida, la Unión Europea (UE) lideran la llamada «carrera» por el desarrollo de la IA en un contexto de confrontación, el desarrollo tecnológico de la IA es mucho más poliédrico y global que el discurso geopolítico de las grandes potencias, con actores y ecosistemas emergentes nada menospreciables. En línea con las llamadas a crear marcos globales de gobernanza de la IA más inclusivos, este artículo explora cuál es el papel del Sur Global en las geopolíticas de la IA. Para ello, analiza los espacios, actores y preocupaciones actualmente invisibilizados en los discursos dominantes sobre la geopolítica de la IA, que son imprescindibles para entender cómo se van a desarrollar los sistemas algorítmicos en los próximos años.

**Palabras clave:** geopolítica, inteligencia artificial (IA), gigantes tecnológicos, regulación, gobernanza, Sur Global, datos, actores, discursos, colonialismo digital, periferias

**Abstract:** *The rise and deployment of artificial intelligence (AI) is having an exponential impact on the process of transformation and realignment of global geopolitics. While the United States, China and to a lesser extent the European Union (EU) are leading the "race" to develop AI in a climate of confrontation, the technological development of AI is much more multifaceted and global than the geopolitical discourse of the major powers, with emerging actors and ecosystems that should not be underestimated. In line with the calls to create more inclusive frameworks of global AI governance, this paper explores what role the Global South plays in the geopolitics of AI. It analyses the spaces, actors and concerns currently overlooked in the dominant discourses on the geopolitics of AI, perspectives that are essential to understand how algorithmic systems are going to develop in the coming years.*

**Key words:** *geopolitics, artificial intelligence (AI), tech giants, regulation, governance, Global South, data, actors, discourses, digital colonialism, peripheries*

Durante la última década, la irrupción y despliegue generalizado de la inteligencia artificial (IA), como revolución tecnológica, está teniendo un impacto exponencial en todo el proceso de transformación y reordenamiento global. La IA puede definirse de múltiples maneras, dependiendo de la aproximación adoptada y del impacto considerado. Shoshana Zuboff (2019), por ejemplo, utiliza el concepto de «machine intelligence» como sinónimo, por considerar que el término incluye tanto el aprendizaje automático como la producción algorítmica *clásica*, junto con muchas operaciones computacionales a las que a menudo se hace referencia como «análisis predictivo» o «inteligencia artificial». Desde la geopolítica, la IA se entiende como una tecnología «habilitadora», el «facilitador definitivo» (Horowitz, 2018) con un impacto todavía difícil de evaluar en la competencia internacional y el equilibrio de poderes globales. Las tecnologías «habilitadoras» se consideran imprescindibles para el desarrollo de multitud de productos y servicios, como es el caso de la tecnología de los semiconductores, la nanotecnología o la IA.

En este contexto, las denominadas tecnologías emergentes –desde la IA hasta las tecnologías cuánticas o las capacidades cibernéticas avanzadas– han tenido un efecto distributivo del poder, pero también un efecto de difusión (Muñiz, 2023). Asistimos a una reconfiguración de equilibrios y dinámicas, así como a nuevas interdependencias y competición entre actores diversos. La revolución de la IA ha pasado de ser una perspectiva de futuro a una realidad configurada por la tensión entre grandes potencias y gigantes tecnológicos. Estamos en los inicios de un «orden mundial generativo» (Cohen, 2023) y su horizonte todavía es incierto. Así, este escenario de competición tecnológica explica el renovado interés por entender la geopolítica de la IA<sup>1</sup>, que se ha manifestado durante la última década en la proliferación de trabajos (véanse, entre otros: Lee, 2018; Lazard, 2023; Csernatoni, 2024; Stango, 2024) que buscan analizar los diferentes factores o aspectos estratégicos que articulan la competencia por desarrollar o adquirir esta tecnología emergente, como la capacidad de desarrollar grandes modelos algorítmicos o de atraer talento, la posesión de determinados materiales estratégicos, o todas las cuestiones relacionadas con la gobernanza o regulación de la IA. La mayoría de estos estudios se centran en el papel de Estados Unidos, China y la Unión Europea (UE), alineándose mayoritariamente con las aproximaciones más tradicionales de la geopolítica alrededor de las grandes potencias globales, mientras se relega a un papel secundario a la periferia donde, a menudo, se manifiesta la competencia por la hegemonía.

Si bien la geopolítica ofrece una forma espacial de pensar el mundo, que organiza a los diferentes actores, elementos y ubicaciones simultáneamente en un tablero de ajedrez

---

1. Aquí la «geopolítica de la IA» debe ser entendida como un cambio sísmico en la estructura y el equilibrio del poder global, que amenaza el estatus de los estados-nación como principales actores geopolíticos (Bremmer y Suleyman, 2023).

global (Ó Tuathail, 1998), en un mundo cada vez más hiperconectado y acelerado por la revolución tecnológica y las redes económicas globalizadas, el orden geopolítico es cada vez más complejo. Por eso, el foco excesivamente occidental de estos análisis puede resultar problemático por varias razones: por un lado, porque invisibiliza realidades que son imprescindibles para entender cómo se van a desarrollar los sistemas algorítmicos en los próximos años y, por el otro, porque en un momento en que se suceden las llamadas a crear marcos globales de gobernanza de la IA más inclusivos, no tener en cuenta estos espacios y preocupaciones «en los márgenes» no hará más que agravar la actual crisis del multilateralismo.

De esta forma, con la voluntad de reflejar la complejidad geopolítica actual, este artículo utiliza el concepto de «geopolíticas de la IA», en plural, para recalcar la multiplicidad de actores e interrelaciones globales que, en la actualidad, estructuran el orden internacional, la diversidad de espacios en competición, y las distintas agendas, valores e impactos que articulan o derivan de esta competición. Partiendo de esta premisa, y con el objetivo de ofrecer un análisis más poliédrico sobre la geopolítica de la IA, el artículo bebe de trabajos que examinan el desarrollo global de la IA desde la perspectiva del Sur Global, para complejizar el entendimiento de cómo opera la competición estratégica por el liderazgo tecnológico global. En otras palabras, se ofrece un análisis de las *otras* geopolíticas de la IA, identificando aquellos espacios, procesos y actores que suelen quedar en los márgenes. En concreto, se propone dar una respuesta a dos preguntas fundamentales: 1) ¿cuál es el papel del Sur Global en las geopolíticas de la IA? y 2) ¿qué espacios, actores y preocupaciones están actualmente invisibilizados en los discursos dominantes sobre la geopolítica de la IA?

Para ello, el trabajo se desarrolla en cuatro partes: en la primera, de naturaleza más conceptual, se justifica la necesidad de hablar de geopolíticas (en plural) de la IA; en la segunda, se ofrece una revisión de la literatura, partiendo de las aproximaciones más convencionales de la geopolítica de la IA para, a continuación, dar algunas coordenadas sobre dónde se sitúan estos discursos más alternativos sobre la geopolítica de la IA; en la tercera, se introduce y problematiza el concepto de Sur Global como marco para visibilizar aquello que las actuales narrativas sobre la geopolítica de la IA ocultan, esto es, otros actores, otros modos de gobernanza, otras comprensiones de qué significa el poder en relación con la IA; y, por último, el artículo se cierra con una cuarta sección que identifica algunos discursos y preocupaciones del Sur Global que no quedan reflejados en los marcos de gobernanza global actual de la IA.

**En un mundo cada vez más hiperconectado y acelerado por la revolución tecnológica, el orden geopolítico es cada vez más complejo. Por eso, el foco excesivamente occidental de los análisis geopolíticos puede resultar problemático, porque invisibiliza realidades –espacios y preocupaciones «en los márgenes»– que son imprescindibles para entender cómo se van a desarrollar los sistemas algorítmicos en los próximos años.**

## De la geopolítica a las geopolíticas

La geopolítica nos habla del ejercicio del poder. De un mundo en que, desde el realismo político (Waltz 1979), los conflictos y las rivalidades ejercen de componentes estructurales del orden global. El concepto de «geopolítica», acuñado originalmente en 1899 por el politólogo sueco Rudolf Kjellen para significar la relación entre geografía y política, ha ido evolucionando a lo largo del siglo xx. Kjellen y otros pensadores imperialistas entendían la geopolítica como parte del conocimiento imperial occidental que se ocupaba de la relación entre la tierra física y la política. El nazismo asoció el término a su objetivo de *Lebensraum* (la búsqueda de más «espacio vital» para la nación alemana) y, hacia el final de la Guerra Fría, la geopolítica revivió como marco para describir la competencia global entre la Unión Soviética y Estados Unidos por la influencia sobre gobiernos y recursos estratégicos a escala global.

Lejos de aproximaciones como las del geógrafo Saul B. Cohen (1964) que, desde principios de la década de 1960, siempre concibió la geopolítica no como una herramienta de conquista y dominación en interés de algunos estados-nación, sino como una herramienta para promover y gestionar el equilibrio entre potencias en el mundo, el concepto ha hecho fortuna como discurso estructurante de la competición entre estados. Tradicionalmente, la geopolítica designaba, desde el realismo clásico, las rivalidades entre grandes potencias, entrelazando política y geografía en un contexto de imperios coloniales y esferas de influencia (Flint, 2006). Sin embargo, el uso y abuso del concepto actual trasciende su significado original. Con el tiempo, el prefijo «geo-» ha ido perdiendo su significado exclusivamente geográfico para convertirse en la partícula que determina la competición por la defensa de los intereses de las grandes potencias aplicable a una gran variedad de sectores y ámbitos, como la geopolítica del ciberespacio, de los recursos naturales o de elementos no espaciales como los semiconductores (Nickel, 2024; Ó Tuathail, 1998).

Este resurgimiento de la geopolítica como instrumento de análisis sirve también para enmarcar un aumento de la competencia por el poder duro en un orden internacional más multipolar. Un mundo donde China y Estados Unidos rivalizan por el liderazgo global; con conflictos territoriales, como la invasión rusa de Ucrania, o con guerras enquistadas en Siria, Yemen o Gaza donde se miden, también de manera interpuesta, (des)equilibrios regionales, en plena crisis de la legalidad internacional (Stephan, 2023) y de fragmentación de la conectividad global; un mundo en competición tanto por el redescubrimiento de la importancia del poder terrestre (con los debates geopolíticos occidentales sobre la Iniciativa china de la Franja y la Ruta [BRI, por sus siglas en inglés]) como por la militarización del espacio exterior.

Asimismo, la geopolítica ha reaparecido con fuerza en el discurso de los líderes políticos y los medios de comunicación occidentales como concepto para compren-

der y analizar los acontecimientos globales. Sin embargo, autores como Carsten Nickel (2024) consideran que este uso de la geopolítica como término para describir nociones de competencia sistémica por el poder, en un sentido neorrealista (Waltz, 1979), está impulsado de manera decisiva por las preocupaciones occidentales sobre su propia competencia con Rusia y China. La idea del *regreso* de la geopolítica se centra en las grandes potencias globales, mientras invisibiliza la periferia donde, a menudo, se manifiesta esta competencia por la hegemonía. Estas periferias, con una capacidad significativamente menor para acumular y proyectar poder, son tratadas como meros «teatros de operaciones» (Png, 2022); sin embargo, desde un punto de vista geopolítico, están lejos de ser irrelevantes, no solo como espacios de interposición. Por eso, en los últimos años, se han sucedido las interpelaciones a tener en cuenta elementos del discurso del Sur Global, que están en tensión con los marcos dominantes. No se trata solo del ascenso vertiginoso de China e India, ni de los indicios de «agotamiento» de la preponderancia de Estados Unidos. La transición es más global y plural: la consolidación de un «ascenso del Sur», que aporta «nuevas constelaciones de poder, que desafían jerarquías y equilibrios tradicionales» (Sanahuja, 2014); la irrupción de un número cada vez mayor de actores no estatales con influencia geopolítica global, y el aumento del número de escenarios de competición geoestratégica.

En esta redistribución del poder global, el espacio digital se ha convertido en una nueva dimensión geopolítica que se desarrolla desde distintos niveles de confrontación: por modelos de digitalización contrapuestos, entre el autoritarismo digital (o «tecnoautoritarismo») y el «tecnoliberalismo» (Sadin, 2016); por la tensión por el poder asimétrico entre gobiernos y grandes oligopolios tecnológicos; así como por la propia competencia entre el puñado de grandes empresas tecnológicas que rivalizan también por su propia influencia geopolítica. Los poderes digitales han transformado el mundo y el entorno global en el que operan los gobiernos; están impulsando la próxima revolución industrial y pueden determinar cómo los países proyectarán su poder económico y militar (Bremmer, 2021). Internet se ha «geopolitizado» (Barbé, 2023). Eric Sadin (2016) habla de la «silicolonización del mundo».

Por todo ello, la complejidad geopolítica obliga a utilizar el concepto en plural para recalcar la multiplicidad de actores e interrelaciones globales que estructuran el orden internacional, la diversidad de espacios en competición, y las distintas agendas, valores e impactos que articulan o derivan de esta competición. También para asegurarnos un análisis más poliédrico.

## Discursos dominantes que estructuran una competición global

La irrupción de la IA está penetrando en todos los niveles de nuestra sociedad, transformando sistemas políticos y servicios, evolucionando modelos económicos y alterando el mercado laboral<sup>2</sup>. La IA está cambiando la recopilación de inteligencia, las vulnerabilidades individuales y colectivas, e incluso la guerra. Por eso, a escala global, asistimos a una competencia de grandes poderes para asegurarse las ventajas económicas, políticas y militares de la IA en un entorno geopolítico cada vez más fragmentado y polarizado (Lazard, 2023).

### Dimensiones estratégicas

La mayoría de los análisis sobre la geopolítica de la IA y de las nuevas tecnologías suele identificar diferentes elementos o dimensiones estratégicas que articulan la competición entre grandes poderes (véanse, entre otros: Lee, 2018; Lazard, 2023; Stango, 2024). Algunos de los elementos que aparecen de modo más recurrente en estos trabajos son los siguientes: a) los datos; b) el *hardware* o infraestructura que hace posible la IA; c) la innovación o capacidad de desarrollar grandes modelos algorítmicos; d) el talento o capital humano; y e) todas las cuestiones en torno a la gobernanza o la regulación de la IA.

A continuación, se analizarán cada una de estas dimensiones:

- a) Los *datos* a menudo se han descrito como «el combustible que alimenta la IA», debido a su importancia en el entrenamiento de algoritmos. Como norma general, cuanto mayor sea el volumen y mejor la calidad de los datos con los que se entrena un algoritmo, mayor será la eficacia del sistema de aprendizaje automático. De aquí que cuestiones técnicas sobre cómo recopilar, organizar, almacenar y hacer accesibles los datos se hayan convertido en estratégicas (Lazard, 2023: 15). Esto es así tanto desde el punto de vista de las empresas, que compiten por desarrollar los modelos algorítmicos más potentes, como desde el de los países que crecientemente buscan cierto control sobre los datos producidos dentro de sus fronteras, fomentando políticas de nacionalismo de datos o soberanía de datos.
- b) En cuanto al *hardware* o *infraestructuras* que hacen posible la IA, podemos incluir distintos elementos en esta categoría. Para empezar, están los chips especializados, como

---

2. Algunos estudios recientes (Acemoglu, 2024) han apuntado que estamos sobreestimando el potencial impacto económico de los nuevos avances en IA.

las unidades de procesamiento gráfico (GPU, por sus siglas en inglés) o las unidades centrales de procesamiento (CPU, por sus siglas en inglés). Son un elemento estratégico porque es muy difícil que un solo país pueda controlar toda la cadena de valor de estos chips. Actualmente, Estados Unidos domina la propiedad intelectual y su diseño, mientras que China es la principal proveedora de materias primas para su producción, como puede ser la silicón, el galio o el germanio. Sin embargo, ambos países dependen de Corea del Sur, Japón y, sobre todo, Taiwán para su fabricación. Otros elementos que podríamos incluir aquí son la infraestructura vinculada a Internet, como los cables marinos, pero también los centros de datos y los superordenadores.

La confrontación tecnológica entre Estados Unidos y China ya se ha traducido en una guerra comercial y de acceso a elementos clave en la industria de semiconductores y para la fabricación de componentes electrónicos.

- c) Otro elemento decisivo es la *capacidad de innovar*, vinculada con la inversión en investigación en las fronteras del conocimiento (*frontier research*) y desarrollo de sistemas algorítmicos. Algunos indicadores importantes aquí para medir la competición entre países incluyen el número de patentes de IA concedidas y el desarrollo de modelos de IA notables o fundacionales<sup>3</sup>. Respecto a las patentes, China ejerce un liderazgo notable, con el 61,1% de todas las patentes globales, seguida muy de lejos por Estados Unidos (20,9%). Aquí los países europeos ocupan una posición muy marginal, con solo el 2,03% del total de patentes (AI Index Report, 2024: 42). Por el contrario, Estados Unidos lidera tanto el ranquin de países con más modelos de IA notables, como el de modelos fundacionales (ibidem). Aunque la práctica totalidad de estos modelos algorítmicos los desarrolla la industria, no puede obviarse el papel que juegan los gobiernos, apoyando aquellas empresas que tienen más capacidad de convertirse en líderes en el sector.
- d) Igualmente, para que haya innovación, más allá de invertir en I+D, es necesario *capital humano o talento*. En otras palabras, resulta necesario atraer y retener un ecosistema de actores que desarrollen los modelos algorítmicos anteriormente mencionados. Varios estudios han demostrado que la oferta de expertos en IA actualmente es inferior a la demanda de este tipo de profesionales (Pallardy, 2024), lo que conlleva una «guerra por la atracción de talento» (The Economist, 2024), en la que los flujos de movilidad del talento global históricamente han beneficiado a los países del Norte Global. El ejemplo

---

3. El *AI Index Report* (2024: 45) hace una distinción entre el término «modelos notables de aprendizaje automático» y «modelos básicos». Los primeros designan modelos particularmente influyentes dentro del ecosistema de IA. El segundo hace referencia a modelos de IA excepcionalmente grandes entrenados en conjuntos de datos masivos, capaces de realizar una multitud de tareas posteriores. Ejemplos de modelos básicos incluyen GPT-4, Claude 3 y Gemini. Cabe destacar, además, que, si bien muchos modelos fundacionales pueden calificarse como modelos notables, no todos los modelos notables son modelos fundacionales.

emblemático es Estados Unidos, donde la mitad de los profesionales de su industria tecnológica nacieron fuera del país (Zeki, 2024). La otra cara de esta moneda es la fuga de talento que experimentan de manera continuada muchos otros países, como Italia, España, Israel o la India (ibídem). Sin embargo, y como se argumentará más adelante, una mirada más plural sobre el mismo concepto de capital humano podría permitir ir más allá de los desarrolladores de modelos algorítmicos, incluyendo también todos aquellos trabajadores, mayoritariamente situados en el Sur Global, que son instrumentales para entrenar a los algoritmos (por ejemplo, mediante etiquetación).

- e) Por último, asistimos también a una *carrera global por regular la IA*: de China al G-7, pasando por la UE (con la aprobación de una pionera Ley de la IA) y el llamamiento del secretario general de Naciones Unidas para el establecimiento de un organismo regulador global de la IA (Bremmer y Suleyman, 2023; Colomina, 2023).

Tabla 1. Elementos y actores clave que estructuran la competición geopolítica por la IA

Componentes geopolíticos	Principales países/regiones	Principales organizaciones
a) <i>Datos</i> Son vitales para la IA, pero es probable que su ventaja se limite a aplicaciones específicas.	China, India, Estados Unidos	Los actores tecnológicos globales más importantes (Meta, Google, ByteDance, Amazon), instituciones financieras
b) <i>Infraestructura física</i> Especialmente los chips y los centros de datos. Hay que tener en cuenta que la energía barata y los abundantes recursos hídricos ayudan a determinar dónde se construye la infraestructura del centro de datos necesaria para el entrenamiento y la inferencia de la IA.	Estados Unidos, Taiwán, Japón, Corea del Sur, Países Bajos	Actores de semiconductores (NVIDIA, TSMC, etc.), empresas de computación en la nube (Amazon Web Services, Azure, Google Cloud, Alibaba Cloud), IBM, Salesforce
c) <i>Modelos algorítmicos / innovación</i>	Estados Unidos, China, Francia, Alemania, Reino Unido	OpenAI, Mistral, Google, Meta, Baidu, Aleph Alpha
d) <i>El talento</i> Es el factor limitante en todos los demás pilares fundamentales.	Estados Unidos, China, Europa, India	Universidades y centros de investigación, Google DeepMind, OpenAI, Anthropic, etc.
e) <i>Regulación</i>	Europa, Estados Unidos, China, Canadá, India, Israel, Japón, Rusia, Singapur, Corea del Sur, EAU	OECD, UE, G-7, G-20, Naciones Unidas, Consejo de Europa

Fuente: Elaboración propia a partir de Lazard (2023: 12).

## **Actores: grandes potencias y corporaciones tecnológicas**

Estados Unidos, China y, en menor medida, la UE lideran la llamada «carrera» por el desarrollo de la IA en un contexto de confrontación. Al mismo tiempo, estas tres potencias promueven modelos de gobernanza tecnológica muy diferenciados, basados en diferencias ideológicas significativas sobre el papel de los mercados, el Estado y los derechos individuales. Estos modelos podrían caracterizarse, a grandes rasgos, como «la autorregulación capitalista de Estados Unidos», «el tecnoautoritarismo de China», y «la regulación integral garantista de la UE». Así, se está dando una confrontación de modelos que tiene considerables ramificaciones geopolíticas, lo que está condicionando sobremanera las iniciativas del resto de actores del sistema internacional.

Anu Bradford (2023), quien acuñó el concepto de «el efecto Bruselas», ha analizado en profundidad las ramificaciones geopolíticas de la existencia de estos tres poderes digitales dominantes, a los que metafóricamente se refiere como «imperios digitales». La autora defiende que, al igual que lo hicieron los regímenes imperialistas históricos, Estados Unidos, China y la UE exportan sus modelos de gobernanza digital a otros países, extendiendo su influencia y dando forma al orden digital global de acuerdo con sus valores. De este modo, los imperios digitales de la era moderna de Internet ejercen el poder no tanto a través del control territorial, sino exportando su tecnología, empresas y marcos regulatorios, influyendo en las normas y valores globales. Dicho de otro modo: este orden se está configurando no solo a través de los diferentes modelos que promueven las grandes potencias, sino también por la dinámica de competencia mutua por ganar influencia. Y aquí resulta útil la distinción que hace Bradford entre la existencia de dos niveles o dimensiones que estructuran la competición global por la IA: las «batallas horizontales» y las «batallas verticales». El primer nivel se produce entre gobiernos, ejemplificada en los conflictos entre Estados Unidos, China y la UE sobre las normas y valores que deben regir la economía digital. Este es el enfoque que adoptan la mayoría de los análisis y trabajos sobre la geopolítica de la IA.

De hecho, tanto el Gobierno de Estados Unidos como el de China reconocen que la innovación tecnológica es una potente fuente de poder, y esta intensa competencia –subyacente entre los dos países por el control del conocimiento y las comunicaciones mediante regulación e influencia institucional, económica y geopolítica– regirá las relaciones internacionales en las próximas décadas (Wu, 2020). En este contexto, la creciente desconfianza entre naciones ha provocado un aumento de las tendencias proteccionistas y de las aspiraciones de «soberanía digital», entendida como la capacidad de un país para controlar su destino digital, incluyendo el control sobre la cadena de suministro de IA, desde los datos hasta el hardware y el software. En definitiva, avanzamos hacia distintas formas de «desacoplamiento tecnológico» con implicaciones directas en los distintos modelos de regulación de los datos y de gobernanza de la IA (Larsen, 2022).

Paralelamente, como apunta Bradford (2023), esta competición horizontal a menudo está determinada (y se articula) también a través de una competición vertical entre los propios gobiernos y las grandes empresas tecnológicas. Si bien los estados han sido los actores principales de las relaciones internacionales durante los últimos casi 400 años, esa supremacía se ha visto alterada por el poder de un reducido grupo de gigantes tecnológicos que rivalizan con ellos por su influencia geopolítica (Bremmer, 2021). La guerra en Ucrania es el ejemplo más claro de cómo estos oligopolios digitales ejercen de actores internacionales privados. Por un lado, se han consolidado como instrumentos imprescindibles del conflicto –recogiendo y compartiendo datos con gobiernos; controlando la información; eliminando cuentas de redes sociales, o actuando como plataformas de movilización y emocionalidad–. Por el otro, Ucrania es hoy un laboratorio de guerra para las *Big Tech* de la inteligencia artificial. Empresas como Palantir, Microsoft, Amazon, Starlink, Google, Clearview AI u otras han colaborado con las Fuerzas Armadas ucranianas para proporcionar tecnologías avanzadas –desde drones a *software* de reconocimiento facial– al Gobierno de Kíev contra la agresión rusa (Csernaton, 2024). Las grandes empresas de IA son, en estos momentos, juez y parte en la transformación del orden internacional.

El desarrollo tecnológico de esta competición se basa en un modelo de negocio construido sobre la extracción, acumulación y procesamiento de datos, que ha permitido a unas pocas empresas acumular una concentración de poder sin precedentes en la historia de la humanidad. El oligopolio de las cinco grandes empresas tecnológicas del Silicon Valley, las llamadas GAFAM –Google (Alphabet), Apple, Facebook (Meta), Amazon y Microsoft–, tiene, conjuntamente, una valoración de capitalización de mercado total superior al PIB combinado de algunos de los países más ricos del mundo (Mirrlees, 2020). Las GAFAM se han convertido en los principales exponentes –aunque no los únicos– del «capitalismo de plataforma», con un modelo de impacto global construido a partir de la desregulación estadounidense. Son la punta de lanza de lo que Shosana Zuboff (2019) llama «capitalismo de vigilancia». Su poder no es solo estructural –sobre la propiedad y los recursos, con capacidad de influencia sobre políticas y regulaciones públicas–, sino también relacional, ya que utiliza tácticas de persuasión y coerción para sus fines. Toda esta concentración de poder, en un entorno de competición geopolítica, de fragmentación y déficit de regulación, plantea interrogantes sobre la naturaleza democrática de gran parte de lo que sucede en el ámbito digital, así como la influencia de estos gigantes empresariales en la toma de decisiones públicas (Bosoer *et al.*, 2022). Se trata de un reordenamiento del poder a nivel global que Ian Bremmer (2021) bautizó como un «momento tecnopolar», marcado por el ejercicio de una cierta forma de «soberanía» por parte de estos gigantes tecnológicos que se expanden más allá del alcance de los reguladores: aportan recursos a la competencia geopolítica; mantienen relaciones exteriores y deben responder ante sus accionistas, empleados, usuarios o anunciantes. Renée DiResta (2024) los ha elevado

a la categoría de «gobernantes invisibles» con capacidad de influencia económica, política y social, y la osadía de crear realidades a medida que controlan el destino de millones de personas.

En el caso de China, además de los distintos modelos de desarrollo y regulación tecnológica que configuran la competición horizontal entre las grandes potencias globales, también el poder de las grandes empresas digitales –entre ellas las más famosas, conocidas como las «BATX» (Baidu, Alibaba, Tencent y Xiaomi)– se ha convertido en un desafío político para el propio Gobierno del país. La tecnología 5G del gigante chino de las telecomunicaciones Huawei ha generado suspicacias, tensiones y acusaciones por parte de gobiernos occidentales. Empresas como Alibaba y Tencent dominan los sistemas de pago, las redes sociales, el comercio electrónico y la logística, además de invertir en proyectos cruciales en la estrategia exterior del Gobierno chino, como la iniciativa de la Ruta de la Seda Digital, que pretende llevar cables submarinos, redes de telecomunicaciones, y ampliar las capacidades de nube en los mercados emergentes involucrados en la BRI (Bremmer, 2021).

Tanta concentración de poder digital genera desigualdad. Voces pesimistas como la de Yuval Noah Harari (2018) advierten de los peligros de crear «dictaduras digitales», construidas con macrodatos, para concentrar el poder en manos de una élite minúscula constituida en un nuevo «tecnofeudalismo» (Varoufakis, 2024).

## Las *otras* geopolíticas de la IA

A estas alturas, establecida la conexión entre la capacidad transformadora de la IA con determinadas agendas políticas, es imprescindible analizar las *otras* geopolíticas de la IA. Aunque el bipolarismo de la rivalidad entre Estados Unidos y China reviste de una importancia innegable, y la UE no quiere quedarse rezagada, el desarrollo tecnológico de la IA es mucho más poliédrico y global que el discurso geopolítico de las grandes potencias, con la participación de otros actores y ecosistemas emergentes nada menospreciables. Desde el concepto de «colonialismo digital» (Kwet, 2019; Couldry y Mejias, 2019; Birhane, 2020; Mejias y Couldry, 2024), que equipara el poder de los monopolios tecnológicos –su impacto en el desarrollo de productos locales y la dependencia que crean del *software* y la infraestructura occidentales– con el colonialismo tradicional, a la supuesta complicidad de la IA con «la supremacía blanca» (Katz, 2020), los discursos alternativos sobre la IA ponen el acento en el modelo de negocio imperante.

Marie-Therese Png (2022) insiste, al respecto, en la importancia de identificar estos espacios y preocupaciones en «los márgenes» y ser conscientes de que, en la era de los algoritmos, el control y la dominación se ejerce a través de mecanismos invisibles, como son los ecosistemas y las infraestructuras digitales (Birhane, 2020).

## El(los) Sur(es) Global(es) y la IA

El Sur es hoy mucho más que un «mero *objeto* de la tensión entre las grandes potencias», ya que también es *sujeto*, «lo que otorga una cierta acción política a sus integrantes» (Menon, 2024). Pero la categoría de «Sur Global» es un concepto geopolítico (no geográfico) que hace referencia a aquellos lugares situados fuera de los centros de poder económico, cultural y político. Se trata de espacios donde, históricamente, se han concentrado las externalidades negativas del capitalismo; una geografía desterritorializada impresa con legados coloniales. Sin embargo, aunque es una noción útil, ampliamente utilizada en círculos académicos y en medios de comunicación en general, los conceptos de «Sur» y «Norte» globales no están exentos de controversia. En este sentido, conviene reconocer las limitaciones que acarrearán estas definiciones para evitar caer en generalizaciones excesivamente simplistas. Además, este es un concepto relacional, que solo puede existir en oposición a un «Norte Global», y los límites entre los dos son confusos, variables, y socialmente contruidos (Roysen *et al.*, 2023).

**Aunque el bipolarismo de la rivalidad entre Estados Unidos y China reviste de una importancia innegable, y la UE no quiere quedarse rezagada, el desarrollo tecnológico de la IA es mucho más poliédrico y global que el discurso geopolítico de las grandes potencias.**

Una de las principales críticas que se le puede hacer al concepto es que omite la heterogeneidad y la incongruencia

interna de las estrategias y los discursos sobre el desarrollo en los diferentes países que asociamos al Sur Global. Y esto también aplica al ámbito de la IA y la gobernanza global digital en sentido amplio. El ejemplo más evidente es el liderazgo del Gobierno y la industria chinas en la carrera global por dominar la IA, y que hace más problemática que nunca la distinción entre el Norte Global y el Sur Global. Pero existen otros muchos ejemplos que ilustran esta heterogeneidad. Basta comparar el peso y las estrategias de dos países del Sur Global como Emiratos Árabes Unidos (EAU) y Kenia en esta carrera tecnológica. El primero aspira a convertirse en un poder regional en el ámbito de la IA, invirtiendo sumas ingentes de capital en desarrollar modelos algorítmicos (Perrigo, 2024). En cambio, el papel de Kenia en la economía política global de la IA está más asociado con trabajadores que entrenan algoritmos en condiciones precarias (*ibidem*, 2023).

Ante esta realidad y, en parte, como una respuesta a estas críticas, ha ido ganando peso la noción de «sures globales» (en plural), que da cuenta de esta heterogeneidad de sistemas políticos, niveles de desarrollo, ideologías e intereses (geo)políticos, así como de las dificultades existentes de coordinación y movilización colectiva entre esta pluralidad de países. Los «sures globales» son una metacategoría que designa un conjunto tanto material como conceptual (Degila y Valle, 2022). Además, el binarismo Norte/Sur no tiene en cuenta a las comunidades más desfavorecidas dentro de los países más ricos, y viceversa. Así, existen *sures* económicos en el norte geográfico, y *nortes* en el sur geográfico.

Con todo, más allá de las críticas que se pueden hacer al concepto de «Sur Global», la realidad es que el discurso dominante sobre la geopolítica de la IA está liderado por los grandes actores de Occidente y, en especial, por Estados Unidos, reforzado por un modelo de gobernanza global digital caracterizado por la colaboración entre múltiples partes interesadas donde, mayoritariamente, son la industria y las comunidades técnicas de los países occidentales las que fijan los estándares. Esta visión otorga a los países del Sur Global un rol más bien periférico, incluso pasivo, o de receptor de normas sobre IA. Desde esta perspectiva, el Sur Global resulta un concepto útil para visibilizar las prioridades, preocupaciones y voces desplazadas de la «Mayoría Global» (Campbell-Stephens, 2021), habitualmente excluidas, silenciadas y marginadas en los discursos hegemónicos.

Esto no quita que, cuando hablamos del Sur Global, haya que hacerlo con la suficiente granularidad como para reconocer que existen también otras posibles lecturas sobre la importancia y la agencia de estos países en el desarrollo o gobernanza de la IA. Por ejemplo, Marie-Therese Png (2022) argumenta que el Sur Global no debe situarse, en absoluto, en la periferia de los desarrollos emergentes de las tecnologías basadas en datos. Al contrario, la autora se opone a estas narrativas y sostiene que el Sur Global está jugando un papel instrumental a la hora de dar forma al desarrollo de sistemas algorítmicos: son una fuente primordial de extracción de datos comerciales a gran escala, mercados laborales para el etiquetado y anotación de datos, y son espacios para la prueba beta de sistemas de IA. Además, son mayoritariamente los países del Sur Global los que proveen los minerales y materiales de tierras raras necesarios para construir las infraestructuras físicas que hacen posibles los productos y los servicios de IA, como los centros de datos, las unidades de procesamiento gráfico, los transistores, las baterías de litio, etc. (Falk *et al.*, 2024).

**El Sur Global no debe situarse en la periferia de los desarrollos emergentes de las tecnologías basadas en datos; al contrario, está jugando un papel instrumental a la hora de dar forma al desarrollo de sistemas algorítmicos.**

## Las otras potencias de la IA

Como ya se ha argumentado, la mayoría de trabajos sobre la geopolítica de la IA se focalizan exclusivamente en analizar la rivalidad y competición entre grandes potencias: Estados Unidos, China y, en menor medida, la UE. Sin embargo, cada vez son más las voces (Ünver y Feldstein, 2022; Lazard, 2023, entre otros) que defienden que la sobre-representación de las grandes potencias en estos estudios acaba conduciendo a análisis inexactos e incompletos sobre la dinámica de adopción global de la IA, lo que no permite dar cuenta de cómo y por qué los países adquieren ciertas tecnologías emergentes.

Además, cuando se trata del desarrollo y la adopción de la IA, algunas potencias medianas, consideradas como «estados bisagra geopolíticos» –a partir del concepto de «geopolitical swing states» de Jared Cohen (2023) o el de «global swing states» de Kliman y Fontaine (2012)–, poseen una capacidad nada menospreciable para dar forma a esta tecnología, tanto en su desarrollo como en su adopción. En este sentido, Cohen (2023) ha identificado un conjunto de capacidades e intereses diferenciados, que distinguen estos actores de otras potencias y que explican su creciente relevancia. Estos son los siguientes: 1) poder económico o regulatorio para dar forma a la innovación y comercialización de tecnología global; 2) posesión de ecosistemas de talento y tecnologías de IA diferenciados; 3) países que albergan empresas líderes mundiales que les dan control sobre puntos críticos de la cadena de suministro de la IA; y, finalmente, 4) estrategias nacionales claras de IA, así como la capacidad y la voluntad de implementarlas y de invertir recursos materiales.

Por ejemplo, países como India, Singapur, Israel o EAU tienen ecosistemas de IA muy dinámicos y punteros, y están realizando inversiones sustanciales en I+D y en aplicaciones comerciales sumamente sólidas (Cohen, 2023). Todos ellos tienen contextos muy propicios que les convierten en mercados cruciales no solo para la exportación de tecnologías de IA, sino también para convertirse en centros mundiales en innovación de estas tecnologías. A ello cabe añadir la emergencia de ecosistemas punteros de industria e innovación en IA y aprendizaje automático en grandes metrópolis africanas como el «Silicon Cape» en Ciudad del Cabo (Sudáfrica), el «Silicon Savannah» en Nairobi (Kenia), el «Sheba Valley» en Addis Abeba (Etiopía), o el «Yabacon Valley» en Lagos (Nigeria).

En cuanto a inversión en innovación, el caso de EAU resulta particularmente relevante. Este país lanzó una ambiciosa estrategia nacional de IA y estableció el primer Ministerio mundial de IA en 2017. Otras iniciativas reseñables incluyen la apuesta por la formación con la creación de la Universidad Mohamed bin Zayed sobre IA o el desarrollo de *modelos de lenguaje extensivo* (o LLM, por sus siglas en inglés) muy sofisticados de código abierto en idioma árabe (como el Falcon y Jais). Este último aspecto reviste de una importancia notable, puesto que, al abrir su tecnología, este país de 10 millones de habitantes se está posicionando como una alternativa y un aliado para aquellos países en desarrollo que no tienen los recursos para construir sus propias herramientas de IA y temen la excesiva dependencia respecto las grandes potencias (Perrigo, 2024). Así, en cierta medida, EAU está jugando tanto la carta de aliado como de competidor de Estados Unidos y China, en una suerte de relación triangular delicada (Clemmensen *et al.*, 2024).

Asimismo, si tenemos en cuenta que posiblemente la «geopolítica de los semiconductores»<sup>4</sup> va a definir el futuro de la IA, Corea del Sur, Japón, Taiwán o los

---

4. Para una mayor comprensión sobre los elementos que estructuran la rivalidad global por la adquisición de chips, véase el trabajo de Miller (2022) sobre «la guerra de los chips», o el de Kotasthane

Países Bajos toman relevancia en tanto que actores estratégicos imprescindibles, de los cuales dependen las grandes potencias mundiales para su propia competitividad tecnológica. Estos países albergan las empresas de diseño y fabricación de semiconductores más importantes del mundo<sup>5</sup>, y están ubicadas en espacios geográficos críticos para las cadenas de suministro globales (p. ej., el Mar de la China Meridional y el Mar de la China Oriental). Ilustra bien la importancia estratégica de estos países la iniciativa «CHIP 4» que busca forjar una alianza entre Estados Unidos y estos tres países asiáticos, con el objetivo no solo de aislar a Beijing, sino también de crear más diversificación en la cadena de suministro y de proteger la propiedad intelectual de estas empresas.

Por otro lado, como defiende Kate Crawford en su *Atlas de la Inteligencia Artificial* (2021), la IA no es ni «artificial» ni «inteligente», en el sentido de que la pretendida «automatización» de esta tecnología se apoya, en realidad, en la explotación laboral de miles de trabajadores y trabajadoras. En otras palabras, ciertamente no existirían modelos algorítmicos potentes sin inversión en I+D, pero tampoco sería posible su desarrollo ni podrían funcionar sin un «trabajo oculto» de recopilación y anotación de datos, externalizado mayoritariamente por las grandes multinacionales occidentales a través de redes laborales complejas. Esta cantidad ingente de trabajo humano que requiere la producción de la IA, que algunos autores han bautizado como «trabajadores fantasma» (Gray y Siddharth, 2019), suele obviarse en los análisis sobre la geopolítica de la IA, invisibilizando la importancia de mucha fuerza laboral de la India, Kenia, Uganda, Nigeria, o Venezuela.

Sobre este punto, resulta pertinente la reflexión que hace Wirth (2024) sobre cómo la existencia de una narrativa dominante que divide los trabajos vinculados a la producción de IA entre alta/baja calificación (que podría simplificarse como ingenieros versus anotadores), ha contribuido a este proceso de invisibilización. Este autor afirma que «los productores de aprendizaje automático (*machine learning*) del Norte Global suelen enmarcar estas dinámicas neocoloniales como una cuestión de ingenieros altamente calificados que delegan el trabajo más servil a trabajadores poco calificados, cuya contribución a menudo permanece sin reconocer». Sin embargo, esta narrativa de «alta versus baja» calificación deja de tener sentido desde el momento en que se reconocen las habilidades específicas que se requieren para estas tareas que, por otro lado, resultan imprescindibles para entre-

---

y Manchi (2023) para una perspectiva de esta cuestión desde el Sur Global.

- De entre estos, destaca especialmente Taiwán. Este pequeño país insular fabrica el 68% de los semiconductores a nivel mundial, y el 90% de los semiconductores de vanguardia necesarios para aplicaciones avanzadas de IA. Así, la Taiwan Semiconductor Manufacturing Company (TSMC) tiene una posición dominante sin rival sobre el mercado global. Igualmente importante es la compañía neerlandesa ASML (Advanced Semiconductor Materials Lithography), único fabricante mundial de máquinas de litografía ultravioleta extrema (EUV), necesarias para los cada vez más sofisticados circuitos integrados en la fabricación de semiconductores (Cohen, 2023: 10).

nar y perfeccionar modelos algorítmicos. Además, más allá de esta distinción, cabe resaltar el papel de productores de talento de algunos países del Sur Global. Sin ir más lejos, los directores ejecutivos de origen indio han conquistado la industria tecnológica de Silicon Valley<sup>6</sup>. Desde esta perspectiva, sin el talento que Estados Unidos extrae de la India, tampoco habría IA mundial por falta de trabajadores en el rango de los «cualificados».

## Gobernanza de la IA, más allá del Norte Global

Como se apuntaba en la primera parte de este artículo, la gobernanza de la IA representa un componente fundamental de cualquier discusión sobre la geopolítica de esta tecnología emergente. De la misma manera que se está dando una carrera global por adoptar y desarrol-

**Se han identificado cuatro preocupaciones habituales del Sur Global que normalmente no quedan reflejadas en los marcos de gobernanza de la IA: 1) la soberanía digital y el monopolio de infraestructuras y reguladores; 2) el extractivismo; 3) las pruebas beta y las incompatibilidades contextuales, y 4) los mercados de datos y los derechos laborales.**

lar la IA, también existe un proceso análogo de carrera global por «regular la IA» (Iyengar, 2023; Galceran-Vercher, 2023; Smuha, 2021). La mayoría de los países del Norte Global y China ya cuentan con regulaciones de diferente índole para gestionar los riesgos y las oportunidades que ofrecen los sistemas algorítmicos y, en los últimos años, también se ha avanzado significativamente en iniciativas de corte multilateral como las impulsadas por la

OCDE, Naciones Unidas o el Consejo de Europa. Pese a todo, las voces y las prioridades del Sur Global siguen infrarrepresentadas o ausentes en estas iniciativas (Okolo, 2023).

Haciéndose eco de esta problemática, Marie-Therese Png (2022) ha identificado cuatro preocupaciones habituales del Sur Global que normalmente no quedan reflejadas en los marcos de gobernanza de la IA: 1) la soberanía digital y el monopolio de infraestructuras y reguladores; 2) el extractivismo; 3) las pruebas beta y las incompatibilidades contextuales y, finalmente; 4) los mercados de datos y los derechos laborales. Ciertamente, podría argüirse que estas preocupaciones no son exclusivas del Sur Global, aunque como se defiende en este artículo, su impacto suele ser más profundo y generalizado en estos países.

La lógica extractiva como patrón sistémico de las economías de datos se despliega tanto en la extracción de información a gran escala, por parte de grandes corporaciones,

---

6. Los casos más emblemáticos son Satya Nadella y Sundar Pichai, CEO de Microsoft y Alphabet, respectivamente. Para más ejemplos, véase: The Economic Times (2024).

como en la obra de mano digital barata, pasando por el uso extractivista de los recursos naturales que necesitan los sistemas de IA para funcionar. Una dimensión fundamental del extractivismo material es el coste ambiental de los sistemas algorítmicos. Esta es una cuestión que, aunque cada vez está más presente en las discusiones y discursos dominantes sobre la gobernanza de la IA, aún ocupa una posición notablemente periférica en la mayoría de ellos<sup>7</sup>. Además, algunas autoras (Falk *et al.*, 2024) han demostrado cómo la huella ecológica de la IA tiene un mayor impacto en los países del Sur Global.

Para argumentarlo, Falk *et al.* (*ibidem*) defienden que la sostenibilidad del desarrollo de la IA se suele examinar poniendo el foco exclusivamente en el uso del agua y energía que requieren los centros de datos, particularmente durante la fase de entrenamiento de algoritmos. Sin embargo, estos análisis pasan por alto los impactos ambientales a lo largo del todo el ciclo de vida de la IA. Así, su huella ambiental también incluye las actividades mineras (como la extracción de litio en Chile), que plantean riesgos sustanciales para la biodiversidad, los recursos hídricos y el bienestar de las comunidades indígenas. En este sentido, los países del Sur Global soportan de manera desproporcionada los costos de extracción de recursos materiales, la degradación ambiental y los riesgos para la salud, mientras que los beneficios económicos del desarrollo de la IA se concentran en las naciones occidentales más ricas. Por ejemplo, si bien países como la República Democrática del Congo, Rwanda y Brasil son fuentes clave de minerales necesarios para la infraestructura de IA, las etapas más rentables de fundición y refinación suelen ocurrir en países de altos ingresos como China y los del Norte Global. Además, el rápido ritmo del avance tecnológico en IA conduce al creciente problema de los desechos electrónicos, que afecta de manera desproporcionada a regiones como África, América del Sur y el Sudeste Asiático, donde la infraestructura de eliminación y reciclaje suele ser inadecuada.

Otro elemento son las pruebas beta y las incompatibilidades contextuales. Las pruebas beta<sup>8</sup> podrían considerarse como una forma de dar continuidad a la práctica colonial de explotación de poblaciones marginalizadas bajo el argumento del progreso científico y tecnológico, esta vez en el contexto de las industrias algorítmicas (Shakir *et al.*, 2020). Se trata de una situación de explotación algorítmica porque normalmente son organizaciones con sede en países occidentales que utilizan países del Sur Global como campos de pruebas, debido a que estos carecen de salvaguardas y regulaciones preexistentes en torno

---

7. Algunas excepciones incluyen la «Artificial Intelligence Environmental Impacts Act» de 2024, una propuesta fallida del partido demócrata de Estados Unidos. A nivel académico, la cuestión de la sostenibilidad de los sistemas algorítmicos también ocupa una posición minoritaria en la literatura especializada, con excepciones notables como Dauvergne (2020) y van Wynsberghe (2021).

8. Las pruebas beta podrían definirse como «la prueba y el ajuste de versiones tempranas de sistemas de *software* para ayudar a identificar problemas en su uso en entornos con usuarios reales y casos de uso» (Shakir *et al.*, 2020: 10).

a los datos y su uso. Además, la mayoría de estas pruebas violaría las leyes de sus países de origen. Algunos ejemplos de esta práctica incluyen el uso de herramientas de predicción electoral en las elecciones de Kenia y Nigeria, por parte de Cambridge Analytica, antes de su implementación en las democracias occidentales (Estados Unidos y Reino Unido).

Del mismo modo, otra práctica a la que apuntan diversos autores (Png, 2022) es la existencia de incompatibilidades contextuales de los sistemas de IA importados. Esto se debe tanto a la falta de conjuntos de datos de entrenamiento específicos de cada país en las economías en desarrollo, como a una representación errónea de las poblaciones y su comportamiento sociocultural o político en los conjuntos de datos de entrenamiento existentes. Lamentablemente, la mayoría de los marcos de gobernanza global para regular la IA actualmente en discusión obvian o prestan una atención muy marginal a la necesidad de capacitar a los países más vulnerables ante este tipo de intervenciones, para que se doten de las debidas protecciones jurídicas e institucionales que eviten estas prácticas comerciales explotadoras.

Finalmente, el último elemento son los mercados laborales de datos y los derechos de los trabajadores. La economía de la IA se sustenta en millones de «trabajadores fantasma» (Gray y Siddharth, 2019) que, en condiciones de extrema precariedad, anotan y clasifican grandes volúmenes de datos para mejorar la visión artificial, el procesamiento del lenguaje natural u otros tipos de algoritmos. Este tipo de trabajos suelen consistir en tareas repetitivas, que deben realizarse con gran precisión, durante largas jornadas laborales y que a menudo acarrear un costo psicológico importante (Spence *et al.*, 2023) para los trabajadores, especialmente para aquellos que se dedican a la moderación de contenidos. Es por ello por lo que suelen ser contratados por plataformas de anotación especializadas (como Microworkers, Samasource, CrowdFlower o Amazon Mechanical Turk) o por plataformas de microtrabajo (como Hive, Scale, o Mighthy AI) en países del Sur Global, donde las leyes de protección laboral son menos estrictas y se puede contratar con facilidad a personas que trabajan muchas horas por salarios bajos y no negociables, casi siempre sin contrato y en condiciones de extrema precariedad.

En los últimos años, diversas organizaciones<sup>9</sup> han denunciado y documentado sobradamente que estas empresas suelen operar a través de prácticas laborales explotadoras, que despojan a los trabajadores de las mínimas condiciones de trabajo seguras y dignas. Como han apuntado varios expertos, estos trabajos afectan desproporcionadamente a quienes son económicamente vulnerables en países con leyes de protección laboral limitadas: en los barrios marginales de Nairobi (como Kibera) o de Calcuta; refugiados que viven en los

---

9. Son un ejemplo de ello la coalición de trabajadores tecnológicos Turkoptikon, sindicatos como UNI Global Union –federación sindical mundial con afiliados en 150 países– o iniciativas de investigación como Fairwork, que evalúa a nivel internacional las condiciones de trabajo en plataformas digitales dentro de la economía de plataformas (*gig economy*).

campos de Dadaab en Kenia o refugiados palestinos de Shatila en Líbano (Jones, 2021); o incluso entre las poblaciones más marginalizadas y excluidas socialmente del Norte Global (Semuels, 2018), como en algunas prisiones de Finlandia (Meaker, 2023).

## Conclusiones

Hasta ahora, el diseño, desarrollo y despliegue de la IA, así como la investigación y gobernanza de sus desafíos asociados, han sido ampliamente moldeados por el Norte Global y el mundo corporativo que marca la pauta de esta transformación. Son, mayoritariamente, la industria y las comunidades técnicas de los países occidentales las que fijan los estándares del desarrollo tecnológico; y tanto el poder político como corporativo occidental lideran el debate sobre el modelo de gobernanza digital, a pesar de que China fue uno de los primeros países en aprobar su propia regulación de la IA. La visión dominante otorga a los países del Sur Global un rol más bien periférico, incluso pasivo, o de receptor de normas sobre IA.

**Es imprescindible incorporar discursos alternativos sobre la gobernanza de la IA y mitigar la distribución desigual de los riesgos de los sistemas de IA, así como superar los patrones de extracción, exclusión o marginalización de un Sur Global que busca un lugar propio en la revolución digital.**

Sin embargo, como muestra este artículo, hay otros elementos importantes de la geopolítica de la IA, que van más allá de las métricas de poder tradicionales asociadas con las grandes potencias y su competición estratégica por el liderazgo tecnológico global: la inversión en innovación en IA, el cultivo y captación de talento, la generación de exportaciones de tecnología en mercados emergentes, el desarrollo de modelos potentes o el establecimiento de estándares globales en IA. Este trabajo visibiliza otros ecosistemas de innovación alternativos y el papel relevante de países que aportan elementos sin los cuales China y Estados Unidos no podrían estar compitiendo (como los semiconductores producidos en Taiwán, Japón o los Países Bajos).

Por eso es imprescindible incorporar discursos alternativos sobre la gobernanza de la IA y mitigar la distribución desigual de los riesgos de los sistemas de IA, así como superar los patrones de extracción, exclusión o marginalización de un Sur Global que busca un lugar propio en la revolución digital. La integración de voces y perspectivas del Sur Global (gobiernos, sociedad civil, industria, academia) en la formación de procesos de gobernanza es solo el primer paso: «la integración debe ir en conjunto con una co-construcción de roles que garanticen que los actores del Sur Global puedan participar también en la co-gobernanza» (Png, 2022). Deben incorporarse las distintas perspectivas de la multiplicidad de actores, agendas, valores e impactos que articulan o derivan de esta competición tecnológica para evitar que la IA afiance la desigualdad global.

## Referencias bibliográficas

- Acemoglu, Daron. «The simple macroeconomics of AI». *NBER Working Paper Series*, n.º 32487, (2024) (en línea) [https://www.nber.org/system/files/working\\_papers/w32487/w32487.pdf](https://www.nber.org/system/files/working_papers/w32487/w32487.pdf)
- AI Index Report. *Stanford Institute for Human-Centered Artificial Intelligence*, (abril 2024) (en línea) [https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI\\_AI-Index-Report-2024.pdf](https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf)
- Barbé, Esther. «Orden en transición y normas en discusión». *Revista CIDOB d'Afers Internacionals*, n.º 134 (2023), p. 21-36.
- Birhane, Abeba. «Algorithmic Colonization of Africa». *SCRIPTed*, vol. 17, n.º 2 (2020) (en línea) <https://script-ed.org/article/algorithmic-colonization-of-africa/>
- Bosoer, Lucía; Cantero Gamito, Marta; Galariotis, Ioannis e Innerarity, Daniel. «Neither fish nor fowl: the challenges of a democratic AI». *STG Policy Papers*, 2022/30 (noviembre de 2022) (en línea) <https://op.europa.eu/publication-detail/-/publication/83f1cb1b-96e2-11ed-b508-01aa75ed71a1>
- Bradford, Anu. *Digital Empires: The Global Battle to Regulate Technology*. Oxford: Oxford University Press, 2023.
- Bremmer, Ian. «The Technopolar Moment. How Digital Powers Will Reshape the Global Order». *Foreign Affairs*, (19 de noviembre de 2021) (en línea) <https://www.foreignaf-fairs.com/articles/world/2021-10-19/ian-bremmer-big-tech-global-order>
- Bremmer, Ian y Suleyman, Mustafa. «The AI Power Paradox». *Foreign Affairs*, (16 de agosto de 2023) (en línea) <https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox>
- Campbell-Stephens, Rosemary M. «Introduction: Global majority decolonising narratives», en: Campbell-Stephens, Rosemary M. *Educational leadership and the global majority: Decolonising narratives*. Londres: Palgrave Macmillan, 2021, p. 1-21.
- Clemmensen, Andrew G.; Redlich, Rebecca y Rumley, Grant. «G42 and the China-UEA-US Triangle». *The Washington Institute for Near East Policy*, PolicyWatch n.º 3854, (2024) (en línea) <https://www.washingtoninstitute.org/policy-analysis/g42-and-china-uae-us-triangle>
- Cohen, Jared. «The generative world order: AI, geopolitics, and power». *Goldman Sachs*, (14 diciembre de 2023) (en línea) <https://www.goldmansachs.com/insights/articles/the-generative-world-order-ai-geopolitics-and-power.html>
- Cohen, Saul B. *Geography and politics in a world divided*. Londres: Methuen, 1964.
- Colomina, Carme. «Una IA ética: la UE y la gobernanza algorítmica». *Opinió CIDOB*, (diciembre de 2023) (en línea) <https://www.cidob.org/ca/publicacions/una-ia-etica-la-ue-y-la-gobernanza-algoritmica>
- Conley, Heather; de Hoop Scheffer, Alexandra; Weber, Gesine; Baranowski, Michal; Kleine-Brockhoff, Thomas; Glaser, Bonnie S.; Small, Andrew; Lesser, Ian O.; McIlhenny, Wi-

- lliam; Mohan, Garima; Connelly, Aaron; Kausch, Kristina; Ishmael, Len; Ünlühisarcıklı, Özgür; Quencez, Martin y Tausendfreund, Rachel. «Alliances in a Shifting Global Order: Rethinking Transatlantic Engagement with Global Swing States». *German Marshall Fund*, (2 de mayo de 2023) (en línea) <https://www.gmfus.org/news/alliances-shifting-global-order-rethinking-transatlantic-engagement-global-swing-states>
- Couldry, Nick y Mejias, Ulises A. *The Cost of Connection: How Data is Colonizing Human Life and Appropriating It for Capitalism*. Redwood: Stanford University Press, 2019.
- Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.
- Csernaton, Raluca. «Charting the Geopolitics and European Governance of Artificial Intelligence». *Carnegie Europe*, (marzo de 2024) (en línea) [https://carnegie-production-assets.s3.amazonaws.com/static/files/Csernaton\\_-\\_Governance\\_AI-1.pdf](https://carnegie-production-assets.s3.amazonaws.com/static/files/Csernaton_-_Governance_AI-1.pdf)
- Dauvergne, Peter. *AI in the Wild: Sustainability in the Age of Artificial Intelligence*. Cambridge, MA: MIT Press, 2020.
- Degila, Dêlidji E. y Valle, Valeria Marina. «Governing Migration for Development from the Global Souths». *International Development Policy / Revue internationale de politique de développement*, n.º 14 (2022) (en línea) <https://journals.openedition.org/poldev/4544>
- DiResta, Renée. *Invisible rulers. The people who turn lies into realities*. Nueva York: PublicAffairs, 2024.
- Falk, Sophie; van Wynsberghe, Aimee y Biber-Freudenberger, Lisa. «The attribution problem of a seemingly intangible industry». *Environmental Challenges*, vol. 16, (2024), p. 1-10.
- Flint, Colin. *Introduction to geopolitics*. Londres y Nueva York: Routledge, 2006.
- Galceran-Vercher, Marta. «Inteligencia artificial y ciudades: la carrera global hacia la regulación de los algoritmos». *Notes internacionals CIDOB*, n.º 286 (febrero de 2023) (en línea) <https://www.cidob.org/publicaciones/inteligencia-artificial-y-ciudades-la-carrera-global-hacia-la-regulacion-de-los>
- Gray, Mary L. y Suri, Siddharth. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Nueva York: Harper Business, 2019.
- Harari, Yuval Noah. *21 Lecciones para el Siglo XXI*. Barcelona: Penguin Random House, 2018.
- Horowitz, Michael C. «Artificial Intelligence, International Competition, and the Balance of Power». *Texas National Security Review*, vol. 1, n.º 3 (2018), p. 36-57.
- Iyengar, Rishi. «The Global Race to Regulate AI. The intelligence may be artificial, but the regulation is real – or might be». *Financial Times*, (5 de mayo de 2023) (en línea) <https://foreignpolicy.com/2023/05/05/eu-ai-act-us-china-regulation-artificial-intelligence-chatgpt/>
- Jones, Phil. «Refugees help power machine learning advances at Microsoft, Facebook and Amazon». *Rest of the World*, (22 de septiembre de 2021) (en línea) <https://restofworld.org/2021/refugees-machine-learning-big-tech/>

- Jorge-Ricart, Raquel y Álvarez-Aragónés, Pau. «The geopolitics of Generative AI: international implications and the role of the European Union». *Real Instituto ElCano*, working paper, (27 de noviembre de 2023) (en línea) <https://www.realinstitutoelcano.org/en/work-document/the-geopolitics-of-generative-ai-international-implications-and-the-role-of-the-european-union/>
- Katz, Yarden. *Artificial Whiteness. Politics and ideology in Artificial Intelligence*. Nueva York: Columbia University Press, 2020.
- Kliman, Daniel M. y Fontaine, Richard. «Global Swing States: Brazil, India, Indonesia, Turkey and the Future of International Order». *German Marshall Fund*, (noviembre de 2012) (en línea) [https://www.gmfus.org/sites/default/files/2023-01/Global-SwingStates\\_KlimanFontaine.pdf](https://www.gmfus.org/sites/default/files/2023-01/Global-SwingStates_KlimanFontaine.pdf)
- Kotasthane, Pranay y Manchi, Abhiram. *When the Chips Are Down: A Deep Dive into a Global Crisis*. Nueva Delhi: Bloomsbury, 2023.
- Kwet, Michael. «Data Colonialism: US Empire and the New Imperialism in the Global South». *Race & Class*, vol. 60, n.º 4 (2019), p. 3-26.
- Larsen, Benjamin C. «The geopolitics of AI and the rise of digital sovereignty». *Brookings*, (8 de diciembre, 2022) (en línea) <https://www.brookings.edu/articles/the-geopolitics-of-ai-and-the-rise-of-digital-sovereignty/>
- Lazard. «Geopolitics of Artificial Intelligence». *Lazard Geopolitical Advisory*, (17 de octubre 2023) (en línea) <https://www.lazard.com/research-insights/the-geopolitics-of-artificial-intelligence/>
- Lee, Kai-Fu. *AI Superpowers: China, Silicon Valley and the New World Order*. Boston: Houghton Mifflin Harcourt, 2018.
- Meaker, Morgan. «These Prisoners Are Training AI». *WIRED*, (11 de septiembre de 2023) (en línea) <https://www.wired.com/story/prisoners-training-ai-finland/>
- Mejias, Ulises A. y Couldry, Nick. *Data Grab: The new Colonialism of Big Tech and how to fight back*. Londres: WH Allen, 2024.
- Menon, Shivshankar. «El Sur global: de objeto a sujeto del orden internacional», en: Fàbregues, Francesc y Farrés, Oriol (coords.) *Anuario Internacional CIDOB 2025*. Barcelona: CIDOB, 2024, p. 19-29.
- Miller, Chris. *Chip War: The Fight for the World's Most Critical Technology*. Nueva York: Simon & Schuster, 2022.
- Mirrlees, Tanner. «Getting at Gafam's "Power" in Society: A Structural-Relational Framework». *HELIOTROPE*, (enero 2020) (en línea) <https://www.heliotropejournal.net/helio/gafams-power-in-society>
- Miyagusuku, Verónica y Aliaga, Vanessa. «La era de la inteligencia artificial generativa: una visión desde el Sur global». *Política Internacional*, n.º 134 (2023). p. 113-127.
- Muñiz, Manuel. «Diplomacia tecnológica para la era digital». *Revista CIDOB d'Afers Internacionals*, n.º 134 (2023), p.91-102.

- Nickel, Carsten. «What do we talk about when we talk about the ‘return’ of geopolitics?». *International Affairs*, vol. 100, n.º 1 (2024), p. 221-239.
- Ó Tuathail, Gearóid. «Introduction. Thinking critically about geopolitics», en: Ó Tuathail, Gearóid; Dalby, Simon y Routledge, Paul (eds.) *The Geopolitics Reader*. Routledge, 1998, p. 1-12.
- Okolo, Chinasa T. «AI in the Global South: Opportunities and Challenges towards more inclusive governance». *Brookings*, (1 de noviembre de 2023) (en línea) <https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance/>
- Pallardy, R. «The AI Skills Gap and How to Address It». *Information week*. (29 de mayo de 2024) (en línea) <https://www.informationweek.com/it-leadership/the-ai-skills-gap-and-how-to-address-it>
- Perrigo, Billy. «Open AI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic». *Time Magazine*, (18 de enero de 2023) (en línea) <https://time.com/6247678/openai-chatgpt-kenya-workers/>
- Perrigo, Billy. «The UAE Is on a Mission to Become an AI Power». *Time Magazine*, (22 de marzo de 2024) (en línea) <https://time.com/6958369/artificial-intelligence-united-arab-emirates/>
- Png, Marie-Therese. «At the Tensions of South and North: Critical Roles of Global South Stakeholders in AI Governance», en: *FACt '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, (2022), p. 1.434-1.445. DOI: <https://doi.org/10.1145/3531146.3533200>
- Royer, Alexandrine. «The urgent need for regulating global ghost work», *Brookings*, (9 de febrero de 2021) (en línea) <https://www.brookings.edu/articles/the-urgent-need-for-regulating-global-ghost-work/>
- Royen, Rebeca; Koehrsen, Jens; Brühwiler, Nadine y Kos, Lasse. «Global South». *SUN – Soziologie und Nachhaltigkeit*, (2023) (en línea) <https://sun-blog.org/global-south-glossar-soziologie-der-nachhaltigkeit/>
- Sadin, Éric. *La silicolonisation du monde*. Le Kremlin-Bicêtre: Ed. L'Échappée, 2016.
- Sanahuja, José Antonio. «Desarrollo global y países emergentes: retos para la política de cooperación de la UE». *Revista CIDOB d'Afers Internacionals*, n.º 108 (2014), p 67-93.
- Samuels, Alana. «The Internet Is Enabling a New Kind of Poorly Paid Hell». *The Atlantic*, (23 de enero de 2018) (en línea) <https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/>
- Shakir, Mohamed; Png, Marie-Therese y Isaac, William. «Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence». *Philosophy & Technology*, vol. 33, (2020), p. 659-684.
- Smuha, Nathalie, A. «From a ‘race to AI’ to a ‘race to AI regulation’: regulatory competition for artificial intelligence». *Law, Innovation and Technology*, vol. 13, n.º 1 (2021), p. 57-84.

- Spence, Ruth; Bifulco, Antonia; Bradbury, Paula; Martellozzo, Elena y DeMarco, Jeffrey. «The psychological impacts of content moderation on content moderators: A qualitative study». *Journal of Psychosocial Research on Cyberspace*, vol. 17, n.º 4 (2023) (en línea) <https://cyberpsychology.eu/article/view/33166>
- Stango, Antonio. «The geopolitical competition between China and the U.S. in new technologies». *Luiss School of Government*, Working Paper Series, SOG-WP12/2024. (mayo de 2024) (en línea) <https://sog.luiss.it/sites/sog.luiss.it/files/stango%20AS%20REV%20The%20geopolitical%20competition%20between%20China%20and%20the%20U.S.%20in%20new%20technologies.pdf>
- Stephan, Paul B. *The World Crisis and International Law: The Knowledge Economy and the Battle for the Future*. Cambridge: Cambridge University Press, 2023.
- The Economist. «The war for AI talent is heating up», (8 de junio de 2024) (en línea) <https://www.economist.com/business/2024/06/08/the-war-for-ai-talent-is-heating-up>
- The Economic Times. «Indians flying high in Silicon Valley. Destination: C-Suite», (10 de agosto de 2024) (en línea) <https://economictimes.indiatimes.com/tech/technology/indians-flying-high-in-silicon-valley-destination-c-suite/articles-how/112411149.cms>
- Toews, Rob «The Geopolitics of AI Chips Will Define the Future of AI». *Forbes*, (7 de mayo de 2023) (en línea) <https://www.forbes.com/sites/robtoews/2023/05/07/the-geopolitics-of-ai-chips-will-define-the-future-of-ai/>
- Ünver, Akın y Feldstein, Steven. «Sources of AI innovation: more than a US-China rivalry». *Transatlantic Policy Quarterly*, (6 de junio de 2022) (en línea) <http://turkish-policy.com/blog/90/sources-of-ai-innovation-more-than-a-u-s-china-rivalry>
- van Wynsberghe, Aimee. «AI for sustainability and the sustainability of AI». *AI Ethics*, vol. 1, (2021), p. 213-218.
- Varoufakis, Yanis. *Tecnofeudalismo. El sigiloso sucesor del capitalismo*. Bilbao: Ed. Deusto, 2024.
- Waltz, Kenneth N. *Theory of International Politics*. Nueva York: McGraw-Hill, 1979.
- Wirth, Assia «Uncovering facial data annotation». *EASST-AS 2024 Amsterdam: Making and Doing Transformation*, (2024) (en línea) <https://nomadit.co.uk/conference/easst-4s2024/paper/84117>
- Wu, Xiangning. «Technology, power, and uncontrolled great power strategic competition between China and the United States». *China International Strategy Review*, n.º 2 (2020), p. 99-119. DOI: <https://doi.org/10.1007/s42533-020-00040-0>
- Zeki. «The State of AI Talent 2024». *Zeki Research Ltd*, (2024) (en línea) <https://zeki-data.com/report/the-state-of-ai-talent-2024/>
- Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Nueva York: Publicaffairs, 2019.

# Acceso a la justicia en tiempos de IA: ¿hacia una justicia *low-cost*?

## Access to justice in times of AI: towards a low-cost justice?

**Marta Cantero Gamito**

Investigadora, Florence School of Transnational Governance, Instituto Universitario Europeo de Florencia; profesora de Derecho de las Tecnologías de la Información, Universidad de Tartu (Estonia). [marta.cantero@eui.eu](mailto:marta.cantero@eui.eu).  
ORCID: <https://orcid.org/0000-0001-8200-3456>

**Cómo citar este artículo:** Cantero Gamito, Marta. «Acceso a la justicia en tiempos de IA: ¿hacia una justicia *low-cost*?». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p.51-71. DOI: [doi.org/10.24241/rcai.2024.138.3.51](https://doi.org/10.24241/rcai.2024.138.3.51)

**Resumen:** Los debates actuales sobre la introducción de la inteligencia artificial (IA) en la administración de justicia, que se reflejan también en la formulación de políticas, se han centrado principalmente en la salvaguarda de la imparcialidad, la transparencia, la explicabilidad, y la supervisión humana en la toma de decisiones, incluyendo las decisiones judiciales. Sin embargo, en paralelo a la cuestión de si la IA puede resolver las crisis de eficiencia del poder judicial, resulta necesario reflexionar sobre la próxima transformación del poder judicial desde la perspectiva del acceso a la justicia. Este artículo problematiza el uso de la IA en los tribunales y defiende la necesidad de evaluar el impacto de la IA en este ámbito desde la perspectiva del acceso a la justicia, en el contexto de la crisis de (acceso a) la justicia en una economía centrada en la inmediatez y el modelo de consumo *low-cost*.

**Palabras clave:** inteligencia artificial (IA), justicia, algoritmos, *low-cost*, discriminación, vulnerabilidad, acceso a la justicia

**Abstract:** Current discussions on the introduction of artificial intelligence (AI) into the courts, also reflected in policymaking, have largely focused on safeguarding fairness, transparency, explainability and human oversight in decision-making, including judicial decisions. However, in parallel with the question of whether AI can solve the judiciary's crises of efficiency, the forthcoming transformation of the judiciary should be reconceptualised from the perspective of access to justice. This paper is an attempt to problematise the use of AI in the courts against the backdrop of the (access to) justice crisis in an economy focused on immediacy and a low-cost consumption model.

**Key words:** artificial intelligence (AI), justice, algorithms, low-cost, discrimination, vulnerability, access to justice

En la era de la transformación digital, el sistema judicial no ha quedado exento de la influencia de la tecnología. La integración de la inteligencia artificial (IA) en la administración de justicia promete una revolución, al ofrecer una justicia más accesible y económica. Sin embargo, esta promesa también viene acompañada de desafíos éticos, económicos y de eficacia que requieren un análisis pormenorizado. Los tribunales de todo el mundo están experimentando *crisis de eficiencia*, relacionadas con una falta de recursos y unos tribunales cada vez más saturados (Gentile, 2024). En este contexto, surge la pregunta sobre el papel que la toma de decisiones automatizada (ADM, por sus siglas en inglés) y la IA pueden desempeñar en la resolución de esta crisis. Este artículo trata de reflexionar sobre la implementación de la IA en los procesos judiciales y sobre si esta puede realmente resolver los problemas del sistema judicial actual o, si por el contrario, agrava las desigualdades y debilita los principios fundamentales de la propia justicia.

En los últimos años, la narrativa en torno a la justicia ha evolucionado desde una concepción monolítica de los procesos judiciales –tradicionalmente delante de un juzgado o tribunal de justicia– hacia un marco más democratizado y accesible, como por ejemplo la mediación o el arbitraje como foros extrajudiciales para la solución de disputas. Este cambio no propone una reducción de la justicia a su mínimo denominador, sino una reinterpretación de la propia justicia, es decir, hacia un paradigma más accesible dentro de las limitaciones y exigencias de la sociedad moderna. En este sentido, es fundamental entender cómo conceptos clave, como el «acceso a la justicia», se redefinen en el contexto digital: acceso efectivo no solo implica que la justicia sea asequible, sino que sea entendible, eficiente y que garantice un trato equitativo, independientemente de las competencias digitales o los recursos socioeconómicos de los usuarios.

De hecho, el acceso a la justicia, entendido como la capacidad de individuos y empresas para resolver problemas legales de manera justa a través de una variedad de servicios jurídicos y mecanismos de resolución de conflictos (OCDE, 2021), también se ha visto profundamente afectado por la digitalización de los tribunales y la introducción de la IA en el sistema judicial. La posibilidad de digitalizar y automatizar la resolución de conflictos promete mejorar la eficiencia del sistema judicial al reducir los tiempos de respuesta y los costes asociados a la administración de justicia. Sin embargo, una mayor eficiencia no debe comprometer la calidad y la atención individualizada que cada caso requiere. Además, cuando los tribunales incorporan automatización algorítmica en sus sistemas, el poder judicial se ve influido por la capacidad regulatoria de la tecnología.

Las experiencias de diversas jurisdicciones en la integración de la tecnología en sus marcos jurídicos y procesos judiciales ponen de relieve la importancia del contexto específico en la aplicación de nuevas soluciones apoyadas en la tecnología (Carvajal, 2020). Aquí, el contexto hace referencia a las particularidades culturales,

legales y socioeconómicas de cada país o región que influyen en cómo se implementan y regulan las tecnologías en el sistema judicial. Por ejemplo, la aceptación pública de la IA, el grado de digitalización del sistema judicial o el acceso de la población a tecnologías digitales varían de un lugar a otro, siendo diferencias que determinan el éxito o fracaso de las iniciativas tecnológicas. Así, lo que funciona en un país puede no ser aplicable directamente en otro sin las adaptaciones necesarias.

Este artículo explora cómo la IA está reconfigurando el acceso a la justicia, analizando tanto las oportunidades como los riesgos asociados en términos de discriminación algorítmica de la justicia la cual podría estar derivando hacia un modelo *low-cost*. Se podría decir que el concepto de justicia *low-cost* surge en la intersección entre eficiencia en la resolución de litigios y su coste. Sin embargo, este enfoque también puede diluir la profundidad del análisis judicial. La simplificación y estandarización de procesos judiciales mediante IA puede llevar a un tratamiento superficial de los casos, afectando la calidad y el rigor con los que se emiten las decisiones. En ese sentido, es importante resaltar, no obstante, que la digitalización ha permitido que algunos conflictos se resuelvan a través de plataformas en línea sin la necesidad de acudir ante un juzgado o tribunal. Además, la resolución de disputas por vías privadas, como la mediación o el arbitraje, permite que muchos casos se resuelvan fuera de los tribunales tradicionales. Aunque esta externalización puede resultar y de hecho está diseñado para que la justicia sea más rápida y accesible, también plantea preocupaciones sobre su conveniencia, ya que no todos los usuarios tienen acceso a la tecnología o entienden cómo usarla correctamente.

Por tanto, partiendo de la creciente digitalización de la administración de justicia y la creciente privatización de la resolución de controversias, que ya no es un poder exclusivo del Estado, este artículo defiende la necesidad de evaluar el impacto de la IA en la justicia desde la perspectiva del *acceso* a esta. En primer lugar, el análisis se centra en el uso de la IA en el sistema judicial como una respuesta orgánica al fenómeno más amplio de digitalización de la administración pública. En segundo lugar, se hace un intento de conceptualizar la justicia *low-cost* en un contexto social y de mercado que demanda soluciones rápidas a los problemas surgidos fundamentalmente de la economía digital. Por último, tras analizar las respuestas del legislador europeo a los problemas planteados por el uso de la IA para la toma de decisiones, se subrayan las limitaciones de un enfoque centrado únicamente en la imparcialidad, la transparencia, la explicabilidad y la supervisión humana en la toma de decisiones judiciales (Kontinen *et al.*, 2022). Este análisis se sitúa en el marco de los debates actuales sobre la eficiencia de la IA frente a la calidad de la justicia y cómo la privatización y algoritmización del sistema judicial plantean nuevos retos para garantizar un acceso a la justicia equitativo para todos los ciudadanos.

## El uso de la IA en el sistema judicial

Cada año, millones de ciudadanos se enfrentan a problemas que requieren la intervención de la justicia. La lista de asuntos es amplia, incluyendo, por ejemplo, disputas que van desde la devolución de bienes de consumo por mal funcionamiento hasta problemas más graves como discrepancias sobre salarios, el riesgo de desahucio, bancarrota o, incluso, negociaciones en torno al cuidado de menores y adultos dependientes. El abanico y la naturaleza de los asuntos ante los juzgados pone de relieve la importancia de contar con sistemas jurídicos accesibles y funcionales para gestionar y mitigar las repercusiones de los problemas que las originan.

Sin embargo, ante un sistema generalmente tensionado, y en muchos casos saturado, los tribunales de todo el mundo afrontan la necesidad de hacer

**El uso de la IA en los tribunales no es únicamente una herramienta para una mayor eficiencia del sistema judicial, sino también una fuente de nuevos desafíos que debe abordar el legislador.**

frente a los asuntos acumulados y garantizar una justicia adecuada (Mora-Sanguinetti, 2022). Según la Comisión Europea para la Eficacia de la Justicia (CEPEJ), el tiempo medio que se tarda en resolver los asuntos civiles y mercantiles en

la Unión Europea es de 237 días, variando entre los estados miembros. Por ejemplo, en Italia, el tiempo estimado que tardan los tribunales en llegar a una decisión en primera instancia es de 674 días<sup>1</sup>, mientras que en Estados Unidos la duración media de los asuntos civiles en los tribunales estatales es de 11,5 meses; y entre 12 y 18 meses en el Reino Unido. Se observan tendencias similares en Canadá o Australia, donde estos asuntos tardan entre seis meses y dos años, dependiendo de la complejidad y la naturaleza del caso. En jurisdicciones como Brasil, el plazo de resolución es de 600 días (Castelliano y Guimaraes, 2023). A los dilatados procesos, debido a la sobrecarga de los tribunales, se suma el hecho de que el número de tribunales en Europa se redujo en un 10% entre 2010 y 2018 (CEPEJ, 2022). Más allá de estas jurisdicciones, estos datos estadísticos reflejan una tendencia mundial y subrayan la necesidad de soluciones innovativas en los procesos judiciales.

Aparte de los tiempos de resolución, otro factor crítico que afecta al acceso a la justicia es el coste. Según un estudio de la OCDE (2021) sobre el acceso a la justicia, las costas procesales constituyen una barrera significativa para muchas

---

1. Datos disponibles en: <https://public.tableau.com/app/profile/cepej/viz/OverviewEN/Overview>

personas, especialmente en casos civiles y mercantiles. Estas costas suponen entre el 10% y el 50% del valor de los litigios, lo que deja a muchas personas y empresas fuera del sistema judicial formal. Esta realidad agrava aún más la desigualdad de acceso, haciendo que el factor económico actúe como un mecanismo de exclusión de facto para los más vulnerables. A lo largo de los años, se han tomado diversas iniciativas, tanto dentro como fuera de los tribunales, para hacer frente a los problemas de acceso a la justicia, como los procesos monitorios o de menor cuantía, o la promoción de mecanismos alternativos (*Alternative Dispute Resolution* [ADR]) y resolución de litigios en línea (*Online Dispute Resolution* [ODR]). Más recientemente, la creciente digitalización de los tribunales y la introducción gradual de herramientas basadas en la IA en el sistema judicial permiten responder a los retos que actualmente afronta la administración de justicia bajo la promesa de reducir drásticamente la duración y el coste de los procesos.

Así, de manera gradual, las herramientas de IA se están utilizando en el sistema judicial para desempeñar diversas funciones dentro del ámbito de la tramitación procesal, la investigación penal e incluso las decisiones judiciales, que pueden variar desde el análisis de grandes cantidades de datos hasta la propia automatización de procesos de toma de decisiones (Lorenzo Pérez, 2022). En un primer nivel, encontramos tecnologías como el archivo electrónico de expedientes (*e-filing*), las videoconferencias y los sistemas de gestión de casos basada en IA, que pueden agilizar ciertas operaciones judiciales, reduciendo el tiempo y los recursos necesarios para tramitar los casos. En un segundo nivel, la IA puede también utilizarse para ayudar en tareas como la transcripción y la traducción, servicios de IA que en los tribunales generan grabaciones instantáneas, lo que mejora la accesibilidad y la eficiencia de los servicios judiciales, y también permiten eliminar las barreras lingüísticas garantizando procedimientos más inclusivos. En un tercer nivel, las herramientas de IA pueden simplificar y agilizar los procesos judiciales siendo usadas como *apoyo* a la identificación de antecedentes e, incluso, para predecir resultados como la duración de la condena y los índices de reincidencia. Por último, en un cuarto nivel se sitúa el uso de la tecnología para la automatización y algoritmización de decisiones judiciales y administrativas. Ello presenta una oportunidad para mejorar el acceso a la justicia, particularmente para personas con recursos limitados o que enfrentan barreras geográficas, ya que facilitan el procesamiento de grandes volúmenes de casos de manera eficiente y reducen tanto los tiempos de espera como los costes. Aquí estaríamos hablando de la creación de un sistema basado en correlaciones estadísticas y el uso de datos masivos para la toma de decisiones sin intervención humana. Por ejemplo, la obtención de un permiso administrativo que autorice la tala de árboles en Estonia se puede hacer a través de un sistema de toma de decisiones automatizada (Pilving, 2023).

No obstante, el uso de la IA en los tribunales no es únicamente una herramienta para una mayor eficiencia del sistema judicial, sino también una fuente de nuevos desafíos que debe abordar el legislador, particularmente en el tercer y cuarto nivel. Los algoritmos suelen ser opacos (*black box problem*), lo que impide la posibilidad de comprender cómo se han generado los resultados (Burrell, 2016 y Castelvechchi, 2016). De esta forma, los algoritmos pueden incluir códigos erróneos y, como resultado, la opacidad algorítmica afecta al derecho a un proceso con las debidas garantías (Palmiotto, 2021). Además, los sistemas de IA se entrenan a partir de datos históricos, que pueden reflejar sesgos y discriminación (Kleinberg *et al.*, 2018). En este sentido, en respuesta a las afirmaciones de que la «algoritmización» de los tribunales aumenta la eficiencia, la literatura sobre la materia se ha centrado especialmente en analizar la cuestión de la imparcialidad, la equidad y la no discriminación (Wachter *et al.*, 2021; Xenidis y Senden, 2020)Mittelstadt, y Russell 2021; Xenidis y Senden 2019. Si no se diseñan y supervisan adecuadamente, los sistemas de IA podrían perpetuar o incluso exacerbar las desigualdades existentes.

En este contexto, en el caso de España, varias iniciativas políticas y legislativas, como la Ley 18/2011, de 5 de julio, reguladora del uso de las tecnologías de la información y la comunicación en la administración de justicia, también «ley de actuación judicial automatizada», o la llamada Ley de Eficiencia Digital del Servicio Público de Justicia<sup>2</sup>, pretenden garantizar que las herramientas de IA utilizadas en el sistema judicial sean transparentes y sujetas a responsabilidad como requisitos indispensables para mantener la integridad del sistema judicial. Ahora bien, aunque se garantice una adjudicación justa y transparente, ¿es posible sostener que la IA solucionará las crisis de (acceso a) la justicia? Existen razones para responder de forma negativa a esta pregunta.

En primer lugar, se puede argumentar que el acceso a la justicia es escaso. Solo un número limitado de asuntos, y algunos tipos de problemas, obtienen una solución por la vía judicial. Por ejemplo, solo el 3% de los litigios de consumo llegan a los tribunales (Comisión Europea, 2023). Esto no significa que los consumidores no obtengan una solución satisfactoria (los datos muestran que el 81% de los consumidores que tienen un problema acuden directamente al comerciante o proveedor de servicios y el 65% de ellos se declaran bastante o muy

---

2. Real Decreto-ley 6/2023, de 19 de diciembre por el que se aprueban medidas urgentes para la ejecución del Plan de Recuperación, Transformación y Resiliencia en materia de servicio público de justicia, función pública, régimen local y mecenazgo.

satisfechos con la forma en que se ha tratado su reclamación), pero el acuerdo alcanzado no siempre se corresponde con la opción de resarcimiento prevista por la ley (ibídem).

En segundo lugar, el recurso a la justicia, entendida como la interposición de una demanda judicial ante juzgados o tribunales, es cada vez menos habitual. En este sentido, los marcos reguladores existentes, como por ejemplo las leyes para las telecomunicaciones, energía y otros sectores regulados, dan prioridad a la tramitación interna de las reclamaciones y a la resolución extrajudicial de los litigios. Esto no es necesariamente objetable en sí mismo. De hecho, la exigencia de establecer mecanismos extrajudiciales de resolución de litigios suele ir acompañada de garantías específicas –procedimientos con bajas costas procesales, transparentes, imparciales, y más rápidos. Por otro lado, además de reducir la brecha en el acceso a la justicia, estos procedimientos también incluyen otras mejoras procesales (Casarosa, 2023). Por ejemplo, en los litigios relativos a la moderación de contenidos por parte de plataformas digitales, las normas procesales contenidas en el Reglamento europeo de Servicios Digitales (DSA, por sus siglas en inglés)<sup>3</sup> podrían incluso suponer una mejora en el ejercicio de los derechos de los usuarios de estos servicios. Este reglamento introduce la obligación de establecer procedimientos claros y transparentes para la revisión de decisiones automatizadas sobre moderación de contenidos por parte de las plataformas digitales, permitiendo a los usuarios apelar de manera más efectiva cuando sus publicaciones han sido eliminadas o restringidas injustamente (art. 15). Además, el DSA obliga a las plataformas a ofrecer explicaciones detalladas sobre las razones de la moderación, lo que aumenta la transparencia y refuerza el derecho de los usuarios a conocer cómo se toman las decisiones que les afectan (art. 17). Asimismo, un importante avance para el acceso a la justicia es la creación de mecanismos de resolución extrajudicial de disputas (art. 21), lo que facilita el acceso a la justicia sin necesidad de recurrir a largos procesos judiciales. Este marco normativo, por tanto, no solo mejora la protección de los derechos digitales, sino que también democratiza el acceso a mecanismos extrajudiciales con debidas garantías a la vez que ofrecen vías más accesibles y rápidas para que los usuarios puedan defender sus derechos frente a las grandes plataformas tecnológicas. Hasta la fecha, se ha publicado información sobre más de 9.000 millones de decisiones de moderación de

---

3. Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE. DO L 277 de 27 de octubre de 2022, p. 1-102.

contenidos, las cuales son a menudo automatizadas, en la base de datos para la transparencia bajo el DSA<sup>4</sup>. Sin embargo, lo anterior puede ser un reflejo del debilitamiento del papel de los juzgados y tribunales como vía principal para obtener resarcimiento, así como un desplazamiento de la relevancia de la justicia como una función de carácter público.

## **(Acceso a la) Justicia *low-cost***

El concepto de justicia de bajo coste o justicia *low-cost* aquí expuesto pretende reflejar la idea de una resolución de disputas más eficiente en términos de tiempo y de coste. Sin embargo, ¿qué noción de justicia opera dentro de la dimensión de la justicia *low-cost*?

El análisis de la propia idea de justicia requiere atender a las condiciones de acceso a la justicia. Para Cappelletti, la propia definición de justicia presupone su acceso efectivo, en términos

**El concepto de justicia de bajo coste o justicia *low-cost* aquí expuesto pretende reflejar la idea de una resolución de disputas más eficiente en términos de tiempo y de coste. Sin embargo, ¿qué noción de justicia opera dentro de la dimensión de la justicia *low-cost*?**

de asequibilidad, disponibilidad y funcionalidad (Garth y Cappelletti, 1978). Sin embargo, el acceso a la justicia, como concepto, no es un proceso estático. Así, en los siglos XVIII y XIX el acceso a la justicia fue entendido como el derecho a litigar o a defender una causa. Como tal,

este derecho se consideraba un derecho natural que no requería un reconocimiento por parte del Estado, aunque sí una garantía frente a su vulneración (ibídem). En este sentido, el acceso a la justicia era un aspecto meramente formal, únicamente al alcance de quienes podían permitírselo; por tanto, el remedio ante la incapacidad de hacer un uso efectivo del derecho y de las instituciones no era tarea del Estado (ibídem: 183). Posteriormente, con las constituciones del siglo pasado, centradas en el Estado del bienestar, el acceso a la justicia se erigió como el requisito fundamental para poder ejercer los derechos consagrados en los nuevos sistemas legales (Galanter, 2010).

En este contexto, la justicia *low cost* puede definirse como un planteamiento de la resolución de controversias que delega las funciones de adjudicación a ins-

---

4. Véase: <https://transparency.dsa.ec.europa.eu/>

tancias extrajudiciales y/o a la tecnología, otorgando prioridad a la asequibilidad y accesibilidad para la ciudadanía. Pero, al igual que otros servicios *low-cost*, este entendimiento de la justicia en términos de eficiencia presenta el posible riesgo de diluir la calidad y el rigor que tradicionalmente se espera del sistema judicial en un Estado de derecho. Si bien el uso de la tecnología puede aumentar las posibilidades de obtener resarcimiento, también suscita preocupación por la falta de profundidad de la tramitación y resolución judicial (Coglianese y Dor, 2021), el grado de atención individual que reciben los casos (Re y Solow-Niederman, 2019), y la posibilidad de que se produzcan errores que puedan afectar a la toma de decisiones (Araujo *et al.*, 2020).

El paradigma de la justicia *low-cost* es en parte tecnocrático y en parte prescriptivo. El modelo de justicia de bajo coste está impulsado por la necesidad de ofrecer un sistema judicial más asequible, especialmente en una época en la que los gastos asociados a procedimientos judiciales pueden resultar prohibitivos para muchos (OECD, 2021). También, en un contexto de creciente deuda pública, puede contribuir a la mejora y, más importante aún, a la democratización del acceso a una tutela judicial efectiva. Sin embargo, esa tutela no está necesariamente encarnada por un órgano judicial o juez(a), como es el caso de los mecanismos de solución alternativa de controversias. Estos mecanismos, como la mediación o el arbitraje, permiten a las partes resolver sus disputas fuera de los tribunales tradicionales, a menudo de manera más rápida y económica. Un ejemplo de esto son las plataformas de resolución de disputas en línea (ODR, por sus siglas en inglés), donde las personas pueden presentar sus quejas, recibir orientación y, en algunos casos, obtener una decisión vinculante sin necesidad de acudir a un juzgado. Más pronto que tarde, estas herramientas se trasladarán a aplicaciones móviles o plataformas digitales que permiten a los usuarios presentar denuncias y recibir decisiones de forma casi instantánea, sin la intervención de un juez humano. Este avance amplía y democratiza el acceso a la justicia, facilitando que más personas, especialmente aquellas que no tienen los recursos para litigar en tribunales formales, puedan resolver sus disputas de manera eficiente (Katsh, 2001). En este sentido, la justicia de bajo coste surge como respuesta a la necesidad de una justicia más accesible económicamente, pero conlleva una tensión implícita entre la accesibilidad y el mantenimiento de los principios fundamentales y, en definitiva, de la propia esencia de la justicia.

Las personas que acuden a los tribunales suelen esperar que sus casos sean oídos y resueltos con un alto grado de atención, conocimientos, imparcialidad y equidad (Comisión Europea, 2023); buscan una decisión justa y que refleje la complejidad de sus circunstancias específicas; esperan, además, competencia en la representación, imparcialidad en el juicio y un proceso que permita la adecuada

presentación y consideración de las pruebas (Bradford *et al.*, 2018). Con la justicia de bajo coste, existe un conflicto subyacente entre tales expectativas y la naturaleza *estandarizada* de los servicios prestados en un proceso digitalizado. En este sentido, las críticas al enfoque neoliberal de la economía proporcionan un paralelismo útil para entender la transformación dentro del sistema judicial. Del mismo modo que las políticas neoliberales ampliaron el alcance del mercado a través de la defensa de una intervención mínima del Estado, lo que llevó a una reconsideración de sus repercusiones sobre la desigualdad y el bienestar público, el desplazamiento hacia soluciones de justicia de bajo coste suscita una reflexión similar. En última instancia, la incorporación de la IA al sistema judicial refleja una aplicación más amplia de la tecnología para abordar necesidades sociales complejas, pero también pone de manifiesto la preocupación por la mercantilización de la justicia (Susskind, 2021) y la erosión de los valores jurídicos centrados en el ser humano (Glover, 2014).

En el intento de definir con mayor precisión y ahondar en el concepto de la justicia *low-cost*, resulta imprescindible abordar el tema desde los múltiples contextos en los que opera la justicia hoy en día. Por ejemplo, las observaciones sobre la Junta de Supervisión de Facebook (FOB, por sus siglas en inglés) proporcionan una perspectiva más actual a través de la cual contemplar el concepto de justicia en la era digital (Klonick, 2018). El FOB representa un experimento de gobernanza privada, que materializa un intento de dirimir los asuntos de moderación de contenidos con un enfoque *cuasi* judicial. Esta Junta fue creada por Facebook para actuar como una especie de *tribunal* supremo independiente, que revisa las decisiones de moderación de contenidos tomadas por la plataforma. Su función es escuchar apelaciones de usuarios cuyas publicaciones han sido eliminadas o restringidas, y emitir fallos que Facebook debe acatar. Lo que diferencia a la FOB de otros mecanismos es que está compuesta por expertos independientes de todo el mundo, confiriéndole legitimidad y autoridad fuera del control directo de la empresa. Además, sus decisiones se publican de manera transparente, proporcionando razones detalladas detrás de cada resolución, lo que contribuye a la claridad en las políticas de moderación. No obstante, esta iniciativa es el reflejo de una tendencia más amplia hacia la privatización de los procesos de toma de decisiones, tradicionalmente competencia de las instituciones públicas, aun cuando dichas decisiones afecten a derechos fundamentales.

De este modo, la privatización de la administración de justicia plantea cuestiones fundamentales sobre las concepciones instrumental e ideológica de la justicia. Desde una perspectiva instrumental, los sistemas privatizados suelen ofrecer una mayor eficiencia, pero carecen de la profundidad y el rigor de los pronunciamientos judiciales que ofrecen los tribunales públicos. Por tanto,

la principal preocupación es que la mercantilización de la justicia pueda conducir a un sistema en el que el proceso debido sea algo secundario frente a la obtención de una decisión inmediata, como requiere el actual modelo económico. Una concepción instrumental de la justicia, centrada en la eficiencia y los resultados del sistema jurídico, pone el acento en los aspectos tangibles de la administración de justicia: la rapidez con que se resuelven los litigios, la accesibilidad del sistema y la eficacia de los tribunales. Esto situaría en primer plano el debate sobre la efectividad de los derechos frente a la mejora de los medios de tutela judicial.

Por el contrario, una concepción ideológica de la justicia se sustenta en los fundamentos filosóficos y morales del derecho. Esta concepción, basada en los derechos, es un intento de interpretar la justicia como exigencia de respeto a la dignidad de todas las personas como agentes morales racionales, libres e iguales. No obstante, la conexión entre justicia e ideología es controvertida. Esto se debe a la diversidad de definiciones de ideología y a las distintas formas en que la ideología puede relacionarse con el derecho. Una relación que, a menudo, se hace con un espíritu crítico, con el fin de impugnar el derecho. La visión del derecho como ideología es una visión que entiende la ideología como fuente de manipulación (por ejemplo, Marx). El derecho como ideología dirige a sus ciudadanos de maneras no evidentes para los propios ciudadanos; el derecho, desde este punto de vista, encubre el poder. El ideal del derecho, por el contrario, implica un conjunto de instituciones que regulan o restringen el poder con referencia a nociones de justicia.

Ambas dimensiones de la justicia forman parte necesariamente de la naturaleza contextual del derecho. La «ficción de la neutralidad» de los sistemas judiciales es especialmente relevante en los debates sobre la IA en la justicia. Aunque los sistemas judiciales aspiran a la imparcialidad, están inevitablemente condicionados por las estructuras sociales y los sesgos que les dan forma (Cappelletti, 1993). Esta interpretación puede complementarse con la «ficción de la competencia», según la cual los conocimientos especializados de los que presumen los sistemas de IA podrían erróneamente percibirse como infalibles o universalmente válidos, con el riesgo de considerar que los individuos pueden reducirse a categorías específicas (Innerarity, 2023). Estas ficciones han de ser objeto de un examen crítico, especialmente a medida que delegamos cada vez más en la tecnología para la impartición de justicia.

**Al igual que otros servicios *low-cost*, este entendimiento de la justicia en términos de eficiencia presenta el posible riesgo de diluir la calidad y el rigor que tradicionalmente se espera del sistema judicial en un Estado de derecho.**

## ¿Democratización o discriminación?

Con el creciente desarrollo de la vida *en línea*, el concepto y las percepciones de justicia también se van expandiendo orgánicamente hacia la dimensión digital. De hecho, ya desde los primeros años de Internet, existían grandes expectativas sobre las posibilidades de las herramientas en línea para reducir la brecha de acceso a la justicia (Casarosa, 2009; Katsh y Rule, 2015). Se pensaba que la digitalización y el uso de plataformas en línea podrían hacer el sistema judicial más accesible, eliminando las barreras físicas y económicas que impedían a muchos ciudadanos acceder a la justicia. La esperanza era que, mediante el uso de sistemas de resolución de disputas en línea, los litigantes pudieran resolver conflictos de manera más rápida y económica, sin los altos costes que

**La idea de que un mayor volumen de datos conduce a mejores decisiones judiciales, aunque atractiva, resulta sumamente compleja. Sin duda, los datos pueden proporcionar información de gran utilidad y mejorar la eficiencia, pero también pueden introducir nuevos sesgos y desigualdades si no se gestionan adecuadamente.**

conlleva la representación legal y sin necesidad de desplazarse hasta un juzgado o tribunal. Además, se creía que la posibilidad de acceder a información legal y asistencia a través de Internet democratizaría el acceso a la justicia, brindando a personas de bajos recursos o en áreas rurales las mismas oportunidades de defender sus derechos que aquellos con más

recursos. En resumen, se esperaba que las herramientas en línea no solo simplificaran los procesos judiciales, sino que también los hicieran más inclusivos y equitativos, en esencia, más *justos*. De ahí que, de la misma manera que la tecnología ha cambiado infinitas profesiones como, por ejemplo, la práctica de la medicina, no resulta descabellado pensar que la IA ha iniciado un proceso de cambio de la práctica del derecho y, por ende, de transformación de la justicia (Barona Vilar, 2021). Además de las herramientas ya existentes, la pandemia de 2020 aceleró la introducción de la tecnología en la administración de justicia, garantizando la continuidad de las funciones judiciales incluso durante el confinamiento (Delgado Martín, 2021).

En perspectiva, el concepto de la justicia del futuro prevé un sistema judicial plenamente integrado con la IA, el *Big Data* y otras tecnologías avanzadas. Como se ha señalado anteriormente, la gestión de los datos puede ayudar en las tareas judiciales tanto de gestión como de decisión. Sin embargo, la idea de que un mayor volumen de datos conduce a mejores decisiones judiciales, aunque atractiva, resulta sumamente compleja. Sin duda, los datos pueden proporcionar información de gran utilidad y mejorar la eficiencia, pero también pueden introducir nuevos sesgos y desigualdades si no se gestionan adecuadamente.

## La consideración del uso de la IA en el sector judicial como sistemas de alto riesgo

Consciente de estos riesgos, en el Reglamento de IA de la UE del 13 de junio de 2024, conocido también como la Ley de IA (AIA, por sus siglas en inglés [AI Act])<sup>5</sup> el legislador europeo clasifica los sistemas de IA utilizados en la administración de justicia como de alto riesgo (véanse artículo 6 y Anexo III). Esta clasificación somete a estos sistemas de alto riesgo a una serie de requisitos concebidos para garantizar su imparcialidad y transparencia. Por ejemplo, la Ley de IA exige la identificación y evaluación de riesgos y la implantación de sistemas de gestión para eliminar o mitigar dichos riesgos (art. 9) con el fin de identificar y mitigar los posibles daños asociados al uso de la IA. Además, los sistemas de IA de alto riesgo deben entrenarse con conjuntos de datos de alta calidad que sean representativos y estén lo más libres de sesgos posible (art. 10). Este requisito aborda la cuestión fundamental de la calidad de los datos, garantizando que los sistemas de IA no refuercen prejuicios históricos. Asimismo, la legislación también exige la documentación y el mantenimiento de registros para asegurar la transparencia y trazabilidad de los sistemas de IA de alto riesgo, incluida información sobre los datos utilizados, el proceso de diseño y desarrollo, y el rendimiento del sistema (art. 11). Los desarrolladores deben mantener una documentación detallada de los datos utilizados, las metodologías empleadas en el entrenamiento de sus sistemas de IA y el rendimiento de estos sistemas. Esta documentación debe ser accesible a los reguladores y, en algunos casos, al público, para aumentar la transparencia y la rendición de cuentas.

Si bien la transparencia hace visible el funcionamiento de los algoritmos de IA, la rendición de cuentas exige que las entidades que aplican estos sistemas respondan de sus decisiones y corrijan los errores. En este sentido, el artículo 22 del Reglamento General de Protección de Datos (RGPD)<sup>6</sup> resulta de interés, ya que reconoce a las personas el derecho a no ser objeto de decisiones basadas

- 
5. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) DOUE L, 2024/1689, 12.7.2024.
  6. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) DOUE L 119 de 4 de mayo de 2016, p. 1/88.

únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos en ellas o les afecten de manera significativa. Esta disposición subraya la importancia de la supervisión humana en los procesos de toma de decisiones basados en la IA, garantizando que las personas tengan derecho a impugnar y solicitar la intervención humana en las decisiones tomadas por los sistemas de IA. La Ley de IA, en su artículo 14, complementa esta disposición haciendo hincapié en la necesidad de una sólida supervisión humana. Los operadores humanos deben tener la capacidad de supervisar los sistemas de IA e intervenir cuando sea necesario para garantizar que las decisiones sigan siendo justas y equitativas. Esta disposición es crucial para mantener la integridad de los procesos judiciales, ya que permite que el juicio humano corrija o anule decisiones potencialmente erróneas de la IA.

## **Los requisitos de la Ley de IA no bastan para prevenir la discriminación porque no solventan el problema de acceso a la justicia**

Los requisitos mencionados anteriormente son un paso acertado hacia la protección de (algunos) derechos fundamentales. Desde esta perspectiva, la transparencia y la explicabilidad conforman la piedra angular del régimen regulatorio de la IA, ya que la información sobre el diseño de los sistemas y los datos usados para su entrenamiento resultan cruciales para garantizar la rendición de cuentas y la equidad de su uso. En este sentido, se necesitan políticas y prácticas proactivas para identificar y mitigar los sesgos en los sistemas de IA, incluidas las auditorías periódicas de algoritmos y la incorporación de diversas perspectivas en el desarrollo de dichos sistemas. Los sistemas de IA transparentes permiten hacer un seguimiento de los procesos de toma de decisiones, lo que facilita tanto a los reguladores como a las personas afectadas comprender cómo se toman las decisiones. Esto es particularmente importante en el sector judicial, donde las decisiones pueden tener profundas repercusiones en la vida de las personas. Sin embargo, a pesar de la importancia de la transparencia y la explicabilidad, existen problemas asociados con la retórica (e incluso fetichismo) de la transparencia, como, por ejemplo, que no produce mejor información, no genera una mayor disciplina de mercado por parte de quienes han de cumplir con la regulación y, por último, tampoco contribuye a la equidad procesal.

En primer lugar, no en todos los supuestos la transparencia y la explicabilidad garantizan que los individuos comprendan el funcionamiento de procesos algorítmicos o, más específicamente, por qué han tomado una determinada

decisión, lo que disminuye las posibilidades de apelar decisiones automatizadas o asistidas por la IA (Edwards y Veale, 2018; Kontiainen *et al.*, 2022). En este sentido, la complejidad de los algoritmos, junto con las tecnologías patentadas y los conocimientos técnicos necesarios para interpretarlos, hacen que la transparencia a menudo no alcance su propósito. Es importante diferenciar entre *explicabilidad* y *comprensión*. La *explicabilidad* de una decisión algorítmica se refiere a la capacidad del sistema para proporcionar información sobre cómo se llegó a una decisión, es decir, a través de qué pasos y criterios. Sin embargo, esto no garantiza que los usuarios, incluidos jueces o abogados, realmente comprendan esa información técnica. La *comprensión* va más allá de la mera explicación, ya que implica la capacidad de interpretar esa explicación de manera significativa y de utilizarla para tomar decisiones informadas o apelar adecuadamente. De esta forma, podemos argumentar que la interpelación a la transparencia y explicabilidad en el uso de la IA para las decisiones judiciales puede crear una falsa sensación de responsabilidad. Incluso si el funcionamiento de un sistema de IA se hace público, ello no se traduce necesariamente en una comprensión o en una toma de control significativa sobre el funcionamiento del algoritmo. Los entresijos técnicos de la IA pueden seguir siendo inaccesibles para la mayoría de las partes interesadas, incluidos jueces, abogados y personas afectadas.

En segundo lugar, la transparencia se ha convertido en un mero ideal normativo invocado tanto por los estados como por las empresas privadas para reforzar su legitimidad (Willett y Morgan-Taylor, 2012). La paradoja de la transparencia sostiene que la relación entre veracidad y legitimidad en las prácticas de transparencia es inherentemente problemática (Koivisto, 2022). La mera exposición del funcionamiento de un sistema de IA no garantiza que esté libre de sesgos o errores. Por el contrario, puede crear una apariencia de legitimidad, especialmente en el caso de actores privados, que oculte y/o perpetúe problemas de fondo (Edwards y Veale, 2018; Zalnieriute, 2021).

Y, en tercer lugar, por ellas mismas, la transparencia y explicabilidad no bastan para prevenir la discriminación o abordar plenamente los problemas de acceso a la justicia. La justicia no solo debe ser transparente, sino también accesible para todos. Esto significa garantizar que todo el mundo,

**A pesar de la importancia de la transparencia y la explicabilidad, existen problemas asociados con la retórica (e incluso fetichismo) de la transparencia, como, por ejemplo, que no produce mejor información, no genera una mayor disciplina de mercado por parte de quienes han de cumplir con la regulación y, por último, tampoco contribuye a la equidad procesal.**

independientemente de su estatus socioeconómico, tenga acceso a una tutela judicial efectiva y a una representación procesal y defensa técnica. La transparencia debe ir acompañada de medidas que promuevan la equidad y la accesibilidad, como políticas que aborden la discriminación algorítmica y garanticen un acceso equitativo a la justicia.

Teniendo en cuenta lo anterior, la adopción de la IA en la administración de justicia exige reflexionar sobre cuestiones clave: ¿Qué concepción de la justicia se persigue promover? ¿Cuáles son las necesidades sociales que cubre el uso de la IA en el sistema judicial? ¿Y cuáles son las consecuencias potenciales, incluidos los riesgos de exclusión en el acceso a la justicia, que se deben gestionar? Las respuestas a estas preguntas conformarán no solo el futuro de la justicia, sino también los contornos del derecho y de nuestro

**La adopción de la IA en la administración de justicia exige reflexionar sobre cuestiones clave: ¿Qué concepción de la justicia se persigue promover? ¿Cuáles son las necesidades sociales que cubre el uso de la IA en el sistema judicial? ¿Y cuáles son las consecuencias potenciales, incluidos los riesgos de exclusión en el acceso a la justicia, que se deben gestionar? El éxito del sistema judicial debe trascender las métricas cuantitativas.**

sistema jurídico. El debate actual sobre la justicia, y el acceso a esta, también debería ampliarse para abordar la creciente delegación, automatización y privatización de la resolución de controversias, como se observa en el DSA y otros marcos legislativos que, como se ha mencionado, fomentan una delegación sistemática hacia la resolución extrajudicial de litigios.

En vista de lo anterior, la justicia *low-cost* se presenta como un subproducto de la actual «economía de la inmediatez», basada en un paradigma de resolución de controversias eficiente en términos de tiempo y costes, y medida por el número de casos *resueltos*. Sin embargo, el éxito del sistema judicial debe trascender las métricas cuantitativas. En su lugar, debe reflejar dimensiones cualitativas que reflejen los valores fundamentales que sustentan el Estado de derecho, lo que incluye un compromiso con el cumplimiento de las garantías procesales y un firme apoyo a la independencia judicial (Gentile, 2022). La búsqueda de la democratización de la justicia (o del acceso a ella), especialmente a través de la IA, debe considerarse con atención para garantizar el fortalecimiento del Estado de derecho (Zalnieriute *et al.*, 2019), que está mostrando signos de debilitamiento en todo el mundo (World Justice Project, 2023).

## Conclusiones: justicia *low-cost* y el futuro del (Estado de) derecho

Los debates sobre el uso de la IA en la toma de decisiones judiciales suelen tener un planteamiento limitado. Al respecto, este artículo ha buscado ampliar el alcance del discurso normativo sobre la democratización y el acceso a la justicia a través de la tecnología. Partiendo de la comprensión de la justicia como un valor fundamental en un Estado de derecho, resulta esencial extender el análisis al impacto de la IA en la calidad y el acceso a la justicia atendiendo a tres puntos esenciales.

En primer lugar, los riesgos asociados a la IA en el poder judicial van más allá de los sesgos e injusticias inmediatos que afectan a grupos específicos. Conviene tener en cuenta repercusiones más amplias, como la forma en que las decisiones algorítmicas pueden crear nuevas barreras para acceder a la justicia. En este sentido, uno de los principales riesgos es el aumento de la brecha digital, ya que los usuarios deben contar con mayores competencias tecnológicas para navegar por sistemas automatizados y acceder a los servicios judiciales en línea. Esto puede excluir a personas de sectores vulnerables o menos familiarizados con la tecnología. Por tanto, deben tomarse medidas prácticas para garantizar que los sistemas de IA se diseñen e implementen de manera que no refuercen inadvertidamente las desigualdades existentes o creen nuevas formas de discriminación en el acceso formal y material a la justicia.

En segundo lugar, aunque la mejora de la eficiencia organizativa de la administración pública a través de la IA es importante, una mayor eficiencia en sí misma no es suficiente para avanzar hacia una verdadera igualdad de oportunidades. La atención no debe centrarse únicamente en hacer que los procesos judiciales sean más rápidos y económicos, sino también en garantizar que los procesos sigan siendo justos y equitativos.

Por último, es necesario considerar también de una manera más amplia el impacto social del despliegue a gran escala de la IA en el poder judicial. Esto incluye el análisis sobre cómo afecta la implementación de la IA a la distribución justa de la riqueza y los recursos, con el potencial de reducir o aumentar las diferencias entre los distintos grupos socioeconómicos. En este caso, el riesgo radica en que la IA se utilice principalmente en sectores desfavorecidos o en casos de menor relevancia, mientras que las decisiones en casos más complejos o vinculados a sectores privilegiados se sigan reservando para jueces humanos. Por ejemplo, la eficiencia de los procesos judiciales gracias a la IA podría reducir los costes, haciendo que la justicia sea más accesible para quienes tienen menos recursos. Sin embargo, sin unas garantías específicas para el acceso a la justicia de

colectividades vulnerables, el uso de la IA podría beneficiar desproporcionadamente a quienes ya cuentan con un mayor acceso tanto a los recursos tecnológicos como de representación legal. De manera que esta dualidad podría aumentar las desigualdades socioeconómicas, reservando la justicia personalizada a aquellos con más recursos y relegando a los menos favorecidos a sistemas automatizados de menor calidad. En este sentido, el despliegue de la IA en los sistemas judiciales no solo debe centrarse en maximizar la eficiencia, sino también en asegurar que se mantenga el acceso equitativo y que no se agraven las desigualdades preexistentes, tanto digitales como socioeconómicas.

En conclusión, la introducción de la tecnología, en particular la IA, en los sistemas judiciales representa una adaptación pragmática a retos contemporáneos de inmediatez y eficiencia. Sin embargo, la externalización de funciones tradicionalmente públicas hacia sistemas algorítmicos puede socavar principios fundamentales de la democracia y erosionar el Estado de derecho. Este enfoque requiere un delicado equilibrio que garantice que los avances tecnológicos no afecten a los derechos y garantías procesales: equidad, imparcialidad, derecho a ser oído y tutela judicial efectiva. De hecho, la dimensión individual de la justicia impone considerar un espectro de factores que no pueden ser abarcados por los planteamientos de «talla única» que, por el contrario, implicarían una estandarización de los procesos judiciales y que, en muchos casos, simplifican en exceso los procedimientos para adaptarlos a un modelo uniforme que no tiene en cuenta las circunstancias particulares de cada litigio o cuestión judicial. Asimismo, si bien la IA puede acelerar los procesos judiciales, reducir costes y hacer que la justicia sea más accesible para un mayor número de personas, también existe el riesgo de que esta eficiencia se logre a expensas de la atención personalizada a cada caso. Por tanto, en la búsqueda de la accesibilidad y la eficiencia, no pongamos en peligro los principios que definen la esencia de la propia justicia.

## Referencias bibliográficas

- Araujo, Theo; Helberger, Natali; Kruikemeier, Sanne y de Vreese, Claes H. «In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence». *AI & Society*, vol. 35, n.º 3 (2020), p. 611-623. DOI: <https://doi.org/10.1007/s00146-019-00931-w>
- Barona Vilar, Silvia. *Algoritmización del derecho y de la justicia: De la inteligencia artificial a la Smart Justice*. Valencia: Tirant lo Blanch, 2021.
- Bradford, Ben, Jackson, Jonathan y Hough, Mike. «Trust in Justice», en: Uslander, Eric (ed.) *The Oxford Handbook of Social and Political Trust*. Oxford University Press, 2016, p. 633-654.

- Burrell, Jenna. «How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms». *Big Data & Society*, vol. 3, n.º 1 (2016) (en línea) <https://doi.org/10.1177/2053951715622512>
- Cappelletti, Mauro. «Alternative Dispute Resolution Processes within the Framework of the World-Wide Access-to-Justice Movement». *The Modern Law Review*, vol. 56, n.º 3 (1993), p. 282-296. DOI: <https://doi.org/10.1111/j.1468-2230.1993.tb02673.x>
- Carvajal, Diana María Ramírez. *Justicia Digital: una mirada internacional en época de crisis*. Universidad de Salamanca, 2020.
- Casarosa, Federica. «Online Auction Sites; An Example of Regulation in Electronic Communities». *European Review of Private Law*, vol. 17, n.º 1 (2009), p. 1-23. DOI: <https://doi.org/10.54648/erpl2009001>
- Casarosa, Federica. «Out-of-Court Dispute Settlement Mechanisms for Failures in Content Moderation Monograph on the Administration of Justice in a Digital Era». *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, vol. 14, n.º 3 (2023), p. 391-402.
- Castelliano, Caio, y Aquino Guimaraes, Tomas. «Court Disposition Time in Brazil and in European Countries». *Revista Direito GV*, vol. 19, (2023), p. 1-24. DOI: <https://doi.org/10.1590/2317-6172202302>
- Castelvecchi, Davide. «Can We Open the Black Box of AI?» *Nature News*, vol. 538, n.º 7623 (5 de octubre de 2016) (en línea) <https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731>
- CEPEJ-European Commission for the Efficiency of Justice. «Special File - Report «European Judicial Systems - CEPEJ Evaluation Report - 2022 Evaluation Cycle (2020 Data)», (5 de octubre de 2022) (en línea) <https://www.coe.int/en/web/cepej/special-file-report-european-judicial-systems-cepej-evaluation-report-2022-evaluation-cycle-2020-data>
- Coglianesi, Cary y Ben-Dor, Lavi M. «AI in Adjudication and Administration». *Brooklyn Law Review*, vol. 86, n.º 3 (2021), p. 791-838.
- Comisión Europea. «Consumer Scoreboard», (27 de marzo de 2023) (en línea) [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_23\\_1891](https://ec.europa.eu/commission/presscorner/detail/en/ip_23_1891)
- Delgado Martín, Joaquín. «Justicia, juicios telemáticos y ecnología para afrontar los efectos de la pandemia sobre la justicia». *Asuntos Constucionales*, vol. 1 (2021), p. 75-94.
- Edwards, Lilian, y Michael Veale. «Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?» *IEEE Security & Privacy*, vol. 16, n.º 3 (2018), p. 46-54. DOI: <https://doi.org/10.1109/MSP.2018.2701152>
- Galanter, Marc. «Access to Justice in a World of Expanding Social Capability». *Fordham Urban Law Journal*, vol. 37, n.º 1 (2010), p. 115-128.

- Garth, Bryant G. y Cappelletti, Mauro. «Access to Justice: The Newest Wave in the Worldwide Movement to Make Rights Effective». *Buffalo Law Review*, vol. 27, (1978), p. 181-292.
- Gentile, Giulia. «AI in the Courtroom and Judicial Independence: An EU Perspective». EUIdeas, (22 de agosto de 2022) (en línea) <https://euideas.eui.eu/2022/08/22/ai-in-the-courtroom-and-judicial-independence-an-eu-perspective/>
- Gentile, Giulia. «Artificial Intelligence and the Crises of Judicial Power: (Not) Cutting the Gordian Knot?», en: De Gregorio, Giovanni; Pollicino, Oreste y Valcke, Peggy (eds.) *Oxford Handbook of Digital Constitutionalism*. Oxford: Oxford University Press, 2024 (en prensa) (en línea) <https://ssrn.com/abstract=4731231>
- Glover, J. Maria. «Disappearing Claims and the Erosion of Substantive Law». *Yale Law Journal*, vol. 124, n.º 8 (2014), p. 3.052-3.093.
- Innerarity, Daniel. «Predicting the past: a philosophical critique of predictive analytics». *IDP. Revista de Internet, Derecho y Política*, n.º 39 (2023), p. 1-12. DOI: <https://doi.org/10.7238/idp.v0i39.409672>
- Katsh, Ethan. «Online dispute resolution: Some lessons from the e-commerce revolution». *Northern Kentucky Law Review*, vol 28, n.º 4 (2001), p. 810-821.
- Katsh, Ethan y Rule, Colin. «What We Know and Need to Know about Online Dispute Resolution White Papers». *South Carolina Law Review*, vol. 67, n.º 2 (2015), p. 329-344.
- Kleinberg, Jon; Ludwig, Jens; Mullainathan, Sendhil y Sunstein, Cass R. «Discrimination in the Age of Algorithms». *Journal of Legal Analysis*, vol. 10 (2018), p. 113-174. DOI: <https://doi.org/10.1093/jla/laz001>
- Klonick, Kate. «The New Governors: The People, Rules, and Processes Governing Online Speech». *Harvard Law Review*, vol. 131, n.º 6 (2018), p. 1.599-1.670.
- Koivisto, Ida. *The Transparency Paradox*. Oxford y Nueva York: Oxford University Press, 2022.
- Konttinen, Laura; Koulu, Riikka y Sankari, Suvi. «Research Agenda for Algorithmic Fairness Studies: Access to Justice Lessons for Interdisciplinary Research». *Frontiers in Artificial Intelligence*, vol. 5, (2022), p. 1-13. DOI: <https://doi.org/10.3389/frai.2022.882134>
- Lorenzo Pérez, Cristina. «Inteligencia artificial en la Administración de Justicia: Regulación española y marco europeo e internacional. Proyectos desarrollados por el Ministerio de Justicia de España». *Centro de Estudios Jurídicos*, (marzo de 2022) (en línea) <https://www.cej-mjusticia.es/sede/publicaciones/ver/13637>

- Mora-Sanguinetti, Juan S. *La factura de la injusticia*. Madrid: Tecnos, 2022.
- OCDE-Organización para la Cooperación y el Desarrollo Económico. «Government at a Glance 2021». *OECD*, (9 de julio de 2021) (en línea) <https://doi.org/10.1787/1c258f55-en>
- Palmiotto, Francesca. «The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings», en: Ebers, Martin y Cantero Gamito, Marta (eds.) *Algorithmic Governance and Governance of Algorithms: Legal and Ethical Challenges*. Cham: Springer, 2021, p. 49-70.
- Pilving, Ivo. «Guidance-Based Algorithms for Automated Decision-Making in Public Administration: The Estonian Perspective». *CERIDAP*, (20 de febrero de 2023) (en línea) <https://doi.org/10.13130/2723-9195/2023-1-105>
- Re, Richard M. y Solow-Niederman, Alicia. «Developing Artificially Intelligent Justice». *Stanford Technology Law Review*, vol. 22, n.º 2 (2019) (en línea) <https://law.stanford.edu/publications/developing-artificially-intelligent-justice-stanford-technology-law-review/>
- Susskind, Richard. *Online Courts and the Future of Justice*. Oxford y Nueva York: Oxford University Press, 2021.
- Wachter, Sandra; Mittelstadt, Brent y Russell, Chris. «Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI». *Computer Law & Security Review*, vol. 41 (2021) (en línea) <https://doi.org/10.1016/j.clsr.2021.105567>
- Willett, Chris y Morgan-Taylor, Martin. «Recognising the limits of transparency in EU consumer law», en: Devenney, James y Kenny, Mel (eds.) *European Consumer Protection: Theory and Practice*. Cambridge: Cambridge University Press, 2012, p. 143-163.
- WJP-World Justice Project. «WJP Rule of Law Index», (2023) (en línea) [Fecha de consulta: 30.06.2024] <https://worldjusticeproject.org/rule-of-law-index>
- Xenidis, Raphaële, y Senden, Linda. «EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination», en: Bernitz, Ulf; Groussot, Xavier; Paju, Jaan y De Vries, Sybe A. (eds.) *General principles of EU law and the EU digital order*. Alphen aan den Rijn: Kluwer Law International, 2020, p. 151-182.
- Zalnieriute, Monika. «“Transparency Washing” in the Digital Age: A Corporate Agenda of Procedural Fetishism Special Issue: Transparency in the Digital Environment». *Critical Analysis of Law: An International & Interdisciplinary Law Review*, vol. 8, n.º 1 (2021), p. 139-153.
- Zalnieriute, Monika; Moses, Lyria B. y Williams, George. «The Rule of Law and Automation of Government Decision-Making». *The Modern Law Review*, vol. 82, n.º 3 (2019), p. 425-455. DOI: <https://doi.org/10.1111/1468-2230.12412>



## ¿Qué dice la música sobre América Latina?

**314**  
NOVIEMBRE-DICIEMBRE 2024

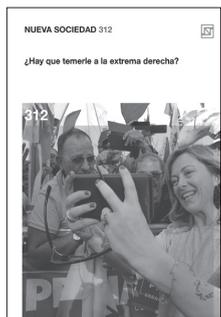
COYUNTURA: **Noam Titelman.** *¿Make Chile Fome Again?* Escenarios preelectorales para 2025.

TRIBUNA GLOBAL: **Gabriel Winant.** El regreso de Trump y las antinomias del Partido Demócrata.

TEMA CENTRAL: **Abel Gilbert.** *El streaming* o la lógica musical del capitalismo financiero. **Mercedes Liska.** Si no puedo perrear, ¿no es mi revolución? Música, sexualización de la cultura y feminismo.

**Florencia Isaura Paparone.** *¿Cómo conquistó el pop coreano América Latina?* Consumos, performances y militancias de los k-popers. **César Jesús Burgos Dávila.** *¿Que truene la tambora y que suene el acordeón!* Composición y consumo de narcocorridos en Sinaloa. **Pablo Schanton.** *Milo J*, el trap argentino y *Los juegos del hambre*. **Rafael González Escalona.** En el Malecón, solo bailamos este son. Cómo el reparto está reinventando la música cubana. **Manuel Soriano.** Cantar, alentar, insultar. De dónde salen los cantitos de cancha. **Pablo Albarces.** *¿El rock resiste?* Roqueros, política y rebeldía en América Latina.

ENSAYO: **Enrique Schmukler.** El sueño eterno del surrealismo.



**312**  
JULIO-AGOSTO 2024

¿Hay que temerle a la extrema derecha?



**313**  
SEPTIEMBRE-OCTUBRE 2024

Guerra, paz y multilateralismo

PAGOS: Solicite precios de suscripción y datos para el pago a <info@nuso.org> o <distribucion@nuso.org>.

EN NUESTRO PRÓXIMO NÚMERO  
**Capitalismo, socialismo y democracia**

# Opini3n p3blica masiva: colonizaci3n algoritmica y *shintetificaci3n* de la esfera p3blica

## Mass public opinion: algorithmic colonisation and *shintetisation* of the public sphere

**Carlos Saura Garc3a**

Investigador predoctoral, Departamento de Filosof3a y Sociolog3a, Universitat Jaume I de Castell3 (Espa3a). *saurac@uji.es*. ORCID: <https://orcid.org/0000-0001-6047-8573>

**Patrici Calvo**

Profesor titular de Filosof3a moral. Departamento de Filosof3a y Sociolog3a. Universitat Jaume I de Castell3 (Espa3a). *calvop@uji.es*. ORCID: <https://orcid.org/0000-0002-3228-9019>

**C3mo citar este art3culo:** Saura Garc3a, Carlos y Calvo, Patrici. «Opini3n p3blica masiva: colonizaci3n algoritmica y *shintetificaci3n* de la esfera p3blica». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 73-95. DOI: doi.org/10.24241/rcai.2024.138.3.73

**Resumen:** Este art3culo profundiza en el cambio estrat3gico de la manipulaci3n de la opini3n p3blica por medio de la colonizaci3n algoritmica, el imperialismo tecnol3gico, la generaci3n de datos y la creaci3n de contenidos sint3ticos. Ello es utilizado por organizaciones gubernamentales, grandes corporaciones tecnol3gicas y poderes econ3micos para amenazar la integridad de la democracia a trav3s de la alteraci3n de los procesos de racionalizaci3n y sentido, as3 como los flujos de comunicaci3n implicados, lo que conduce a la aparici3n de patolog3as sociales, anomal3as y distorsiones en la esfera p3blica digital. El objetivo de este art3culo es mostrar las principales estrategias vinculadas con las tecnolog3as digitales disruptivas para la creaci3n de sentido y la manipulaci3n de la opini3n p3blica y, especialmente, criticar los impactos y desaf3os que subyacen a este nuevo contexto democr3tico *algoritmizado*, *shintetificado* y masivo para la propia opini3n p3blica y la democracia.

**Palabras clave:** *shintetificaci3n*, opini3n p3blica masiva, esfera p3blica digital, algosfera, tiran3a de la mayor3a, inteligencia artificial (IA) generativa, democracia

**Abstract:** *This paper examines the strategic change in the manipulation of public opinion through algorithmic colonisation, technological imperialism, data generation and the creation of synthetic content. This is utilised by government organisations, big tech corporations and economic powers to threaten the integrity of democracy via the alteration of the processes of rationalisation and meaning-making, as well as the communication flows involved, which leads to the appearance of social pathologies, anomalies and distortions in the digital public sphere. The aim of this paper is to show the main strategies related to disruptive digital technologies for the creation of meaning and the manipulation of public opinion and, especially, critique the impacts and challenges for public opinion and democracy underlying this new algorithmised, shintetified and mass democratic context.*

**Key words:** *shintetisation, mass public opinion, digital public sphere, algosphere, tyranny of the majority, generative artificial intelligence (AI), democracy*

## Introducción: de la opinión mayoritaria a la opinión masiva

La dependencia de la ciudadanía de las sociedades modernas de multitud de plataformas digitales (X, Facebook, Instagram, TikTok, Metaverso, etc.) y el desarrollo de los fenómenos de la hiperconectividad, la algoritmización y la datificación han originado que la esfera pública digital aumente su tamaño hasta ocupar prácticamente la totalidad de los espacios de la esfera pública (Hagen *et al.*, 2017). Este acoplamiento e hibridación de ambas esferas está originando grandes cambios en los procesos de construcción de la opinión pública y en el propio sistema democrático (Habermas, 2023). Con la consolidación de la segunda era de la inteligencia artificial (IA), la opinión pública se ha convertido en objeto de deseo de organizaciones gubernamentales, grandes corporaciones tecnológicas y poderes económicos alrededor del mundo (García-Marzá y Calvo, 2022 y 2024; Woolley, 2023).

Estos colectivos han realizado una construcción de la opinión pública de diversas formas. Por una parte, a través de los mecanismos de funcionamiento de la esfera pública digital fuertemente vinculados con las influencias sociales, se ha intentado imponer un espiral del silencio (Noelle-Neumann, 2010) que potencie que las opiniones de los individuos se acerquen y se alineen alrededor de la opinión mayoritaria o, en caso de que no sea así, que el valor, la importancia y el peso de estas opiniones en la esfera pública digital sea silenciado, reducido o eliminado (Sunstein, 2017 y 2019). Por otra parte, se han desarrollado instrumentos e infraestructuras de introducción de datos y contenidos sintéticos y difusión artificial de estos en la esfera pública digital para generar ruido, alterar el funcionamiento de los algoritmos de las plataformas digitales, distorsionar de forma artificial la opinión pública<sup>1</sup> y crear una «opinión pública sintética» (Saura García y Calvo, 2024).

En la actualidad, empero, la disruptiva e implacable colonización algorítmica de la esfera pública y de la propia democracia ha sufrido un giro estratégico. Ya no se trata tanto de inundar todo el ciberespacio con datos y contenidos sintéticos para influir y manipular los procesos de formación de la opinión pública, así como potenciar el desarrollo de un patrón totalmente ficticio creado de forma artificial en favor de los intereses de determinados

---

1. Véanse: Woolley y Howard (2018); Howard (2020); Pérez Zafrilla (2021); Ottaviani (2022); Woolley (2023) y García-Marzá y Calvo (2024).

colectivos que introducen los propios datos y contenidos (García-Marzá y Calvo, 2022; Woolley, 2023); sino que ahora se actúa por medio de diversos instrumentos e infraestructuras en dos niveles distintos: uno subsistémico (en la esfera pública digital), en el cual se introducen aquellos datos que permitan la construcción del patrón de opinión pretendido, y otro sistémico (en la esfera privada y gubernamental), en el cual se detectan exclusivamente los patrones de opinión de la esfera pública digital deseados<sup>2</sup>.

En este contexto, el objetivo de este artículo teórico-práctico es detallar las diversas estrategias, vinculadas con las tecnologías digitales disruptivas, para la creación de sentido de forma artificial y sintética y para la manipulación de la opinión pública, así como ofrecer una crítica de los impactos y consecuencias que esto puede tener para la sociedad y los sistemas democráticos<sup>3</sup>. Para ello, en primer lugar, partiendo de las reflexiones de pensadores como Alexis de Tocqueville y John Stuart Mill, se expondrá el concepto de opinión pública mayoritaria y su influencia en la nueva esfera pública digital, y los peligros que esta implica para la democracia. En segundo lugar, se detallarán las principales características de la «opinión pública sintética» promovida por las tecnologías de la segunda era de la IA. En tercer lugar, se profundizará en el concepto de «opinión pública masiva» y, finalmente, desde una perspectiva ético-crítica, se valorarán peligros actuales y potenciales que subyacen a la nueva situación contextual generada por la irrupción de la citada «opinión pública masiva», tanto para la democracia como para la propia opinión pública<sup>4</sup>.

- 
2. Se entiende por sistema el conjunto de esferas, como el Estado y el mercado, que aplica procesos de racionalización con respecto a fines, mientras que por subsistema se entiende el conjunto de ámbitos del mundo de la vida –objetivo, social y personal– que aplica procesos de racionalización con respecto a sentido (Habermas, 1987).
  3. Para el desempeño de este objetivo, el estudio se ha apoyado en un enfoque metodológico hermenéutico crítico. Siendo propio de la filosofía moral y política (Habermas, 1984), este se aleja de aquellas posturas mecanicistas y científicas basadas en la neutralidad y la objetividad de los datos y metadatos observables y cuantificables, a fin de abordar el interés del conocimiento por una mayor emancipación y autonomía del ser humano a través de la comprensión, aprehensión y crítica de la realidad y justicia social.
  4. Cabe destacar que la actual situación de acelerada evolución y desarrollo de las tecnologías digitales disruptivas y de la IA generativa analizadas en este estudio pueden llegar a producir, por una parte, limitaciones en la actualización del análisis teórico realizado en el texto y, por la otra, transformaciones o modificaciones de los procedimientos, fenómenos, situaciones e impactos analizados.

## La tiranía de la mayoría en la esfera pública digital

El exponencial proceso de hiperconectividad, dataficación y algoritmización de todos los ámbitos, tareas y funciones de las democracias en las sociedades modernas, entendidos en sentido amplio (político, económico y social), está potenciando los peligros que amenazan la existencia de las propias democracias ya enunciados por Alexis de Tocqueville en la primera mitad del siglo XIX y por John Stuart Mill en la segunda mitad del mismo siglo. Entre las citadas amenazas destacan especialmente las vinculadas con la opinión pública, como son la omnipresencia de la opinión mayoritaria, la opresión ideológica de las masas, los efectos negativos de la conformidad, la imposición del silencio, las revoluciones mudas, etc. (Mill, [1861]1972; Tocqueville, [1835-1840]1988). Tanto Tocqueville primero como Mill después subrayan los graves problemas que puede causar en el correcto funcionamiento de la democracia una tiranía de la mayoría, haciendo referencia a la dominación de la opinión pública por la llamada opinión mayoritaria.

**El exponencial proceso de hiperconectividad, dataficación y algoritmización de todos los ámbitos, tareas y funciones de las democracias en las sociedades modernas, entendidos en sentido amplio (político, económico y social), está potenciando los peligros que amenazan la existencia de las propias democracias ya enunciados por Alexis de Tocqueville y John Stuart Mill en el siglo XIX.**

En su análisis de la democracia de los Estados Unidos en el siglo XIX, Tocqueville argumentaba que: «(...) la mayoría tiene un inmenso poder de hecho y un poder de opinión casi igual de grande, y una vez que ha decidido respecto a una cuestión no hay, por así decir, obstáculos que puedan, no diré que detener, sino ni siquiera retrasar su marcha y dejarle tiempo para escuchar las quejas de aquellos que aplasta al pasar» (ibídem: 244).

La dominación de la mayoría expuesta por Tocqueville supone una emergencia en relación con el libre pensamiento y reflexión, y da lugar a una situación en la cual el poder de la mayoría restringe la propia libertad de espíritu y la libertad de discusión de la ciudadanía provocando la desaparición de la crítica y la posibilidad de cambio dentro de la democracia. La tiranía de la mayoría predomina sobre el sentido común, el buen juicio y la propia justicia, lo que provoca que los equilibrios subóptimos se impongan en los procesos de construcción del saber moral y que la reproducción, la repetición y el silencio se adueñen de los espacios de deliberación y diálogo (ibídem). Tocqueville resume estos efectos nocivos diciendo que: «la mayoría traza un círculo formidable alrededor del pensamiento. Dentro de esos límites, el escritor es libre. Pero pobre de él si se atreve a salir de ellos» (ibídem: 250).

Los efectos de la tiranía de la mayoría provocan la marginación de la crítica, el atrofiamiento de la opinión pública y la progresiva homogeneización de esta dentro de una opinión mayoritaria conformada a partir de un conjunto de ideas y acciones intuitivas y automatizadas que han colonizado las formas de pensar, actuar, sentir y vivir de los ciudadanos. En relación con este tema, tanto Tocqueville como Mill argumentan que las mayorías no tienen necesidad de poder político para imponer su dominación social. Por una parte, Tocqueville sostiene que «estar en desacuerdo con la masa equivale, por decirlo así, a no vivir. Esta no tiene necesidad de emplear las leyes para doblegar a los que no piensan como ella, le basta con desaprobarnos. El sentimiento de su aislamiento y su impotencia les abruma y desespera» (ibídem: 307). Por otra parte, Mill subraya el mismo problema cuando afirma que la sociedad misma «(...) practica una tiranía social más formidable que muchos tipos de opresión política, ya que, aunque generalmente no se sustenta con penas tan extremas, deja menos vías de escape, penetra mucho más profundamente en los detalles de la vida y esclaviza el alma misma» (Mill, 1977 [1859]: 220)<sup>5</sup>.

Para Tocqueville y Mill, por consiguiente, a la masa le es suficiente con su poder, basado en la sensación de legitimación por la mayoría social, para imponer su homogeneidad intelectual y comportamental y aplicar correctivos disuasorios sobre las minorías. Es decir, la opinión mayoritaria se expande impulsada por una masa de personas que impone una espiral del silencio en la cual todo lo que sea o parezca diferente a la opinión mayoritaria y toda persona que piense o parezca diferente es excluida, silenciada y reducida (Noelle-Neumann, 2010). De ahí que tanto Tocqueville (1988 [1835-1840]) como Mill (1972 [1861]) alerten de la emergencia de las prácticas despóticas y tiránicas de la opinión mayoritaria, de sus impactos en el funcionamiento de los sistemas democráticos y de la necesidad de hacer frente a estas prácticas y evitar la colonización de la opinión pública y la democracia por parte de una opinión mayoritaria caracterizada por ser agregativa y robusta a cualquier tipo de crítica y a la cual todos los ciudadanos se adhieren y ninguno se atreve a criticar.

En el contexto actual de esfera pública digital, los impactos negativos de las prácticas despóticas y tiránicas de la opinión mayoritaria sobre la opinión pública y los sistemas democráticos expuestos por Tocqueville y Mill se ven representados y potenciados por diversos fenómenos vinculados con las influencias sociales como, por ejemplo, la conformidad social y las cascadas de opinión (Sunstein, 2017 y 2019; Aral, 2021). Por una parte, la promoción de conformidad en

---

5. Nota de Ed.: Esta cita y las siguientes cuya referencia original está en una lengua diferente del castellano son una traducción libre del original.

ecosistemas digitales tiene que ver, principalmente, con la influencia directa de las creencias y los comportamientos de otras personas. Ello se visualiza claramente en el papel que juegan las cámaras de resonancia de las plataformas digitales basadas en mecanismos algorítmicos (Pariser, 2011; Sunstein, 2017; Lanier, 2018; Williams, 2021). Por otra parte, las cascadas de opinión en la esfera pública digital se originan como consecuencias de las llamadas influencias informacionales y reputacionales y hacen referencias a los conjuntos de personas que disminuyen su crítica y aceptan una opinión o una idea debido a las creencias o acciones transmitidas por determinadas noticias o personas (Sunstein, 2017 y 2019; Aral, 2021).

De esta forma, los fenómenos de la cohesión y las cascadas de opinión de grupo en la esfera pública digital acentúan los problemas vinculados con el despotismo y la tiranía de la opinión mayoritaria expuestos por Tocqueville y Mill en el siglo XIX, pero esta vez en el siglo XXI. En un contexto social totalmente digitalizado como el actual, dichos efectos producen un gran impacto negativo en la opinión pública y los sistemas democráticos. Sin embargo, estos representan una sola parte del proceso de instrumentalización de la esfera pública digital para influenciar, controlar, manipular la opinión pública. En la actualidad, las grandes corporaciones digitales, las grandes organizaciones gubernamentales y los poderes económicos han desarrollado y perfeccionado otra forma de aplicar la tiranía y el despotismo mucho más sutil, eficiente y eficaz a través de mecanismos tecnológicos. Esta forma busca aprovechar la generación y difusión de contenidos sintéticos, la alteración del funcionamiento de los algoritmos de las plataformas digitales y la vigilancia social masiva con el objetivo de potenciar la distorsión de los procesos de construcción de opinión pública mediante su *sintetificación*; es decir, a través de verter sobre ellos conjuntos de datos sintéticos capaces de generar «opinión pública sintética» (Saura García y Calvo, 2024; García-Marzá y Calvo, 2024).

## **La construcción de una «opinión pública sintética»**

El constante desarrollo de las tecnologías de la información y la comunicación (TIC) en la primera década del siglo XXI fue esencial para la constitución de una «sociedad red» basada en las redes de información y comunicación digital (Castells, 2006). Esta sociedad red potenció y produjo multitud de cambios políticos, económicos y sociales entre los que destacan la hibridación entre los medios de comunicación tradicionales y los medios de comunicación digitales, el incremento de la interacción de la ciudadanía por medio de la dinamización

y multiplicación de las formas de comunicación o la transformación y digitalización de la opinión y la esfera pública (Papacharissi, 2002 y 2010; Carpentier, 2011; Chadwick, 2013).

La evolución de la nueva situación política, económica y social propiciada por el desarrollo de las tecnologías digitales y, especialmente, las plataformas sociales produjo, en un primer momento, un incremento de las posibilidades de participación democrática de la ciudadanía y un empoderamiento de esta (Papacharissi, 2010; Tufekci, 2017). Sin embargo, con el paso del tiempo, el avance de la hiperconectividad, dataficación y algoritmización de todos los ámbitos, tareas y funciones de las democracias en las sociedades modernas ha propiciado que esta situación cambie y se haya revertido, lo que ha dado lugar a impactos negativos para la soberanía, autonomía y autodeterminación de la ciudadanía, deformándose la libre e igual participación democrática y concentrándose la esfera pública digital en un pequeño grupo de plataformas y servicios digitales propiedad de grandes corporaciones digitales. Ello ha socavado las condiciones básicas para el correcto funcionamiento de los sistemas democráticos (Saura García, 2022 y 2024; Habermas, 2023; García-Marzá y Calvo, 2024).

Las características de las plataformas digitales gestionadas por las grandes corporaciones tecnológicas del mundo han convertido la esfera pública digital en un espacio de hibridación donde convergen acciones, actividades, contenidos y opiniones procedentes tanto de actores humanos como no humanos (García-Marzá y Calvo, 2022 y 2024; Millière, 2022). Este hecho está posibilitando la construcción de una «opinión pública sintética» basada en la introducción de grandes cantidades de contenidos de forma artificial, en la creación de una infraestructura de vigilancia social masiva y en la instrumentalización de la llamada *algoresfera*<sup>6</sup>. Ante esta situación, las amenazas para la democracia descritas por Tocqueville y Mill se acrecientan y se potencian por medio de nuevos fenómenos como el despotismo tecnológico, el individualismo digital y la tiranía algorítmica.

La «opinión pública sintética» está constituida, en mayor medida, por grandes conjuntos de contenidos basados en datos sintéticos generados, vertidos y esparcidos por usuarios, instrumentos y mecanismos artificiales totalmente computarizados, automatizados y *algoritmizados* y, en la mayoría de los casos, controlados, financiados y dirigidos parcial o totalmente por las grandes corporaciones digitales, las grandes organizaciones gubernamentales y los poderes fácticos. En menor

---

6. La *algoresfera* hace referencia al ecosistema ciberfísico controlado y gobernado por algoritmos que permite la hiperconectividad digital, la vigilancia masiva y la dataficación comportamental (García-Marzá y Calvo, 2022).

medida, aquella está constituida por los contenidos naturales introducidos por la ciudadanía, es decir, por «autores autóctonos», que utilizan la esfera pública digital para expresar y discutir sus opiniones con el objetivo de ponerse de acuerdo sobre diferentes cuestiones de este mundo, y por «públicos usufructuarios», que la utilizan para lograr sus intereses e instrumentalizar a la ciudadanía (Habermas, 1998 y 2023). Pero el objetivo primario de la «opinión pública sintética» no es adular la opinión pública por medio de contenidos naturales creados por personas reales, sino inundar la esfera pública digital con datos y contenidos creados artificialmente

**El objetivo primario de la «opinión pública sintética» no es adular la opinión pública por medio de contenidos naturales creados por personas reales, sino inundar la esfera pública digital con datos y contenidos creados artificialmente con el propósito de generar realidad y sentido de forma totalmente artificial.**

con el propósito de generar realidad y sentido de forma totalmente artificial en y sobre cualquier tema, así como de crear una opinión pública maciza que no se pueda ver afectada o alterada por ningún tipo de crítica.

Sobre este punto, cabe destacar que la capacidad de generar cantidades ingentes de contenidos sintéticos y la hibridación de estos con los

creados por humanos provocan que las personas sean incapaces de detectar la *sintetización* de los contenidos e interpreten todos los contenidos de la esfera pública digital como naturales, atribuyéndolos directamente a contenidos creados por personas. A este hecho se le suma la imposibilidad en muchos casos de identificar y verificar los contenidos engañosos o directamente falsos introducidos deliberadamente por «públicos usufructuarios» con la intención de persuadir, manipular o adular la opinión pública, es decir, mediante *fake news*, desinformación, propaganda computacional, *deep fakes*, entre otros (Habermas, 2023; Meikle, 2023; Woolley, 2023) y, ahora, a través de *deep-synthetic* (Calvo y Saura García, 2024).

La fabricación de la «opinión pública sintética» se basa, principalmente, en la creación, inundación y propagación de contenidos creados artificialmente por la IA generativa en la esfera pública digital. Sin embargo, no hay que pasar por alto que también utiliza, por una parte, los mecanismos de los fenómenos de influencia social de las plataformas digitales analizados en la primera sección y, por la otra, la propagación de contenidos personalizados a cada usuario humano con el objetivo de maximizar la generación de realidad y sentido a fin de potenciar los efectos de influencia, control, persuasión y manipulación interesada de la opinión pública de forma artificial. Los tres pilares fundamentales sobre los que se construye la «opinión pública sintética» son a) la intoxicación informativa, b) la colonización algorítmica y la instrumentalización de la esfera pública digital, y c) la vigilancia social masiva.

- a) La *intoxicación informativa* hace referencia al uso por parte de gobiernos, grandes corporaciones digitales, partidos políticos, personajes multimillonarios o «públicos usufructuarios» en general de las nuevas tecnologías digitales disruptivas para adular y manipular la opinión pública por medio de la utilización, principalmente, de *bots sociales* (Howard, 2020; Ottaviani, 2022; Woolley, 2023). Los *bots sociales* hacen referencia a cuentas de las plataformas sociales controladas sobre todo por algoritmos dedicados a crear, introducir y propagar contenidos artificiales para interactuar con usuarios o con otros *bots* y alterar la opinión pública de forma específica (Chang *et al.*, 2021; Hajli *et al.*, 2022). La fuerte irrupción en noviembre de 2022 de la IA generativa y especialmente de los modelos de lenguaje extensivo (LLM, por sus siglas en inglés) ha revolucionado el funcionamiento y los impactos de los *bots sociales* en las plataformas digitales (García-Marzá y Calvo, 2024). Por una parte, ha potenciado la inundación de contenido sintético y la propagación de *fake news*, desinformación o propaganda computacional en la esfera pública digital (Howard, 2020; Woolley, 2023) y, por la otra, ha aumentado exponencialmente la calidad, la minuciosidad y la personalización de los contenidos falsos y de los llamados *deep fakes* (Woolley, 2020; Meikle, 2023); finalmente, ha producido un nuevo fenómeno de pervisión de la opinión pública: las *deep-synthetic* (Calvo y Saura García, 2024), es decir, aquella información que, sin ocultar que se trata de un conjunto de datos y metadatos sintéticos generados por IA generativa, es capaz de influir de forma más profunda y eficiente en la voluntad libre de la ciudadanía gracias al patrón comportamental que le subyace.
- b) La *colonización algorítmica* y la *instrumentalización de la esfera pública digital* han provocado la transformación de esta última en una *algoesfera* controlada por las grandes corporaciones tecnológicas o por poderes fácticos, personajes multimillonarios o gobiernos (Bartlett, 2018; Moore, 2018; Da Empoli, 2020). Por una parte, la mayoría de los algoritmos que componen esta *algoesfera* están gestionados por las grandes corporaciones digitales del planeta —que a su vez son las creadoras y dominadoras del capitalismo de la vigilancia— con el objetivo de satisfacer sus intereses económicos (Zuboff, 2020); y, por la otra parte, dichos poderes fácticos, personajes multimillonarios o gobiernos realizan campañas de alteración y manipulación de los algoritmos de las plataformas sociales a partir de la utilización de las API (*Application Programming Interface*). Las API hacen posible la intercomunicación entre diversas aplicaciones y programas, originándose la alteración y manipulación interesada del funcionamiento de los algoritmos de plataformas, como Facebook, Instagram, X o Google, principalmente por medio de la recreación de movimientos o contenidos artificiales o sintéticos que engañen a los algoritmos (Aral, 2021; Woolley, 2023).

c) La *vigilancia social masiva*, por su parte, también es ejercida por las grandes corporaciones digitales y gobiernos o poderes públicos (Jiang y Fu, 2018; Snowden, 2019; Ebeling, 2022). La infraestructura de la vigilancia social masiva existente ha dado lugar a un panóptico digital que hace posible observar, *dataficar* y explotar todos los clics, los movimientos, las reacciones, las opiniones, etc., dentro de la esfera pública digital y los ecosistemas ciberfísicos (Zuboff, 2020), así como vigilar y *dataficar* a través de dispositivos invasivos o no invasivos los espacios íntimos de las personas al objeto de explotar grandes conjuntos de datos íntimos relacionados con aspectos corporales, sentimentales, desiderativos o mentales (Farahany, 2023).

Los efectos conjuntos de estos aspectos han creado una opinión pública sintética que se mantiene inmóvil y robusta, que no se ve afectada por ningún tipo de crítica, que quiebra y destruye la opinión pública y que está construida, controlada y dirigida por grandes corporaciones digitales, gobiernos, poderes públicos y otros poderes fácticos (García-Marzá y Calvo, 2024). Al respecto, destacan las siguientes consecuencias:

- *Rotura de la acción comunicativa*. La desintegración de la acción comunicativa y la personalización de contenidos desembocan en un incremento del individualismo digital, la destrucción de los espacios discursivos y la potenciación del autoadoctrinamiento (Habermas, 2023).
- *Dataficación de la opinión*. No se realiza una reflexión y valoración de las manifestaciones, acciones u opiniones de las personas, sino solamente un cálculo computacional de los datos y metadatos de conjuntos de usuarios –humanos o no humanos– en relación con diversos temas (García-Marzá y Calvo, 2022).
- *Monopolización de la esfera pública*. La esfera pública digital –y los grandes conjuntos de datos que emanan de ella– está gestionada, controlada, mercantilizada e instrumentalizada por las grandes corporaciones digitales (Saura García, 2022).
- *Preeminencia de los contenidos sintéticos y artificiales*. El uso de las API y de los *bots sociales* con IA generativa ha propiciado la inundación de contenidos sintéticos y la propagación de *fake news*, *deep fakes*, desinformación o propaganda computacional automatizada (Woolley, 2020 y 2023).
- *Dificultad para diferenciar entre lo sintético y lo real*. La imposibilidad de distinguir entre las acciones, actividades, contenidos u opiniones procedentes de actores humanos y las producidas por actores no humanos provoca que las personas en primer lugar los identifiquen todos como realizados por actores humanos (Woolley, 2020 y 2023).

- *Vigilancia ubicua*. Existencia de una infraestructura que monitoriza, vigila y *datafica* constantemente la esfera pública digital y también los espacios privados e íntimos de la ciudadanía (Zuboff, 2020; Farahany, 2023).

Las citadas consecuencias ponen de manifiesto y agravan los principales problemas que se producen para el correcto funcionamiento y la buena salud de las democracias enunciados por Tocqueville y Mill, por diversas razones. En primer lugar, potencian un desequilibrio de los principios de libertad e igualdad de la ciudadanía, necesarios para el correcto funcionamiento de la opinión pública y de la democracia. En segundo lugar, promocionan prácticas nocivas para el correcto funcionamiento de la opinión pública y la democracia, como son el despotismo tecnológico, la tiranía algorítmica y el individualismo digital. En tercer lugar, promueven una colonización de la opinión pública real a través de una «opinión pública sintética», que es impermeable y monolítica ante la crítica creada de forma artificial por poderes públicos, económicos o fácticos basada en la omnipresencia de su opinión, la opresión ideológica de las masas y la imposición del silencio.

Recientemente, la dominación disruptiva de la esfera pública digital, la opinión pública y la propia democracia han sufrido un giro estratégico y, en cierto modo, inesperado. Ya no se trata solo de inundar todo el ciberespacio con datos masivos para influir y manipular. Como se ha mencionado anteriormente, ahora se actúa en dos niveles: uno sistémico y otro subsistémico. Por un lado, uno que opera en la esfera pública digital (subsistema) –detectando y proporcionando aquellos datos que permitan la construcción del patrón deseado– y otro que opera en la esfera privada y gubernamental (sistema) –captando específicamente los patrones de opinión de la esfera pública digital deseados–.

## «Opinión pública masiva»: el impacto de los políticos algorítmicos

Fruto de los últimos avances tecnológicos en procesamiento del lenguaje natural ((PLN, por sus siglas en inglés), especialmente en el desarrollo y uso de la IA generativa y algoritmos estocásticos –como el ChatGPT (*Generative Pre-trained Transformer*) de Open IA, el Gemini de Google, el Llama de Meta o el Claude de Anthropic–, desde 2017 los llamados políticos algorítmicos o virtuales han ido ganando peso, fuerza y presencia en los procesos políticos de elección de representantes, toma de decisiones, deliberación y resolución de conflictos,

generación de programas electorales, construcción de opinión política y sentido social, así como de desarrollo de políticas públicas.

Este proceso comenzó su andadura en 2017 con SAM y Alisa, los primeros intentos de presentar un algoritmo de IA a unas elecciones presidenciales de un país democrático (Nueva Zelanda y Rusia, respectivamente). Continuó su desarrollo en 2019 con IA-Mayor, el primer algoritmo de IA que consiguió participar de facto en unas elecciones municipales (Tama New Town, Tokio, Japón), y en 2022, con Leaders Lars, el primer partido político (*Det Syntetiske Parti*) compuesto íntegramente por un algoritmo de IA que participó en unas elecciones presidenciales (Dinamarca). Y se consagró en 2023, con ION, la primera IA nombrada oficialmente como consejera presidencial de un Gobierno democrático, en Rumania (García-Marzá y Calvo, 2024)<sup>7</sup>. El caso ION es especialmente significativo para la democracia, ya que, dado su reconocimiento público y oficialidad gubernamental, marcó un punto de inflexión en la aplicación e implementación de un enfoque algorítmico de democracia: primero, porque ION es el primer ente no-humano que ha sido nombrado oficialmente como consejero de un Gobierno democrático (Rumanía); segundo, porque la oficialidad de ION no se esconde detrás de un subterfugio legal, un alter ego humano u organizacional como en los casos de Alisa, SAM, IA-Mayor y Leaders Lars; y, tercero, porque sus recomendaciones e informes, aunque no son vinculantes, tienen línea directa con el máximo representante gubernamental.

Tal y como lo describió el Gobierno rumano en su presentación pública, ION –que significa *nosotros* en rumano– es una IA basada en lenguaje extensivo (LLM) que recopila información directa de la ciudadanía a través de su web oficial, e indirecta mediante el rastreo de las redes sociales con el objetivo de conocer el sentir mayoritario y las preferencias e intereses de la ciudadanía rumana digitalmente hiperconectada. A través de ello, responde preguntas, eleva informes técnicos y asesora al presidente del Gobierno rumano sobre un tema concreto: «ION –la aplicación desarrollada por los mejores investigadores y profesores rumanos en este campo– representará el eslabón perdido de la comunicación en tiempo real y en el siglo XXI entre ciudadanos y gobiernos, un sistema que utilizará la IA, el procesamiento de lenguaje natural y la visión por ordenador para captar de forma rápida y automática las opiniones de los rumanos utilizando los datos disponibles públicamente en las redes sociales. (...) ION será la versión moderna de la antigua

---

7. Los últimos casos de políticos virtuales han sido VIC (*Virtual Integrated Citizen*), un chatbot basado en ChatGPT que se ha postulado como alcalde de una población de Wyoming, en Estados Unidos (Elliott, 2024a), y AI Steven, un avatar de IA que se ha postulado como parlamentario del Reino Unido (Elliott, 2024b), ambos en 2024.

ágora, un vínculo entre los ciudadanos y los responsables de la toma de decisiones, donde la voz de todos podrá ser escuchada» (ION, s/f: 1).

En el informe técnico del proyecto ION, disponible en la web del Gobierno rumano, se describen y justifican los dos objetivos principales del desarrollo e implementación práctica del proyecto (ibídem: 9):

- *Participación directa de la ciudadanía en la toma de decisiones política.* ION identifica, recopila, procesa y agrupa la información más relevante de las redes sociales para identificar los principales intereses de la ciudadanía. Para ello, en primer lugar, sondea las redes en busca de aquella información vinculada con el tema a tratar que ha recibido más *likes* y comentarios (relevancia). En segundo lugar, analiza la información recopilada a través de métodos de PLN y aprendizaje profundo. Finalmente, utiliza técnicas y métodos de *data mining* como el *clustering* para agrupar la información y determinar los intereses y las preferencias de la ciudadanía
- *Seguimiento continuo de las repercusiones, los resultados y las dinámicas.* ION, una vez determinados y difundidos entre los políticos los intereses de la ciudadanía alrededor de un tema concreto, se preocupa de realizar un seguimiento continuado de los resultados que estos van teniendo en las políticas públicas y su repercusión práctica, así como sobre la evolución que tales intereses tienen para evitar anacronismos que aumenten la brecha entre los gobernantes y los ciudadanos.

Más allá de lo anecdótico del caso, ION simboliza la conquista de la última frontera de la *algoritmización* de la democracia: la oficialidad de los políticos virtuales. Si hasta 2023 la participación democrática de los políticos virtuales se había logrado a través de un subterfugio legal, amparándose en una identidad humana –Michihito Matsuda y Leader Lars–, con el reconocimiento oficial de ION como cargo público se han roto no solo barreras legales, sino también las psicológicas, sociales y culturales. Además, ION ha producido la emergencia de dos fenómenos subyacentes: la hegemonía de la opinión masiva y el reacoplamiento del mundo sistémico (Estado y mercado) y el mundo de la vida (objetivo, intersubjetivo y subjetivo) a través de la expansión de la *algoresfera* y su hibridación e influencia profunda en la esfera política. La «opinión pública masiva», consiste en la reducción a datos computables y patrones comportamentales de todos los resultados que producen cada uno de las tres formas de opinión (mayoritaria, artificial/sintética y pública) que se producen en la esfera pública para, desde un criterio binario, procesarlos, analizarlos y convertirlos en información relevante y conocimiento aplicable para la toma de decisiones, la resolución de conflictos o la elaboración y desarrollo de leyes o políticas públicas; es decir, para conocer y aplicar el sentido de algo en la democracia.

A diferencia de los *bots sociales* tradicionales o generativos, que actúan en la esfera pública digital promoviendo o produciendo datos y metadatos sobre algo de la democracia, ION actúa en la esfera política extrayendo patrones procedentes de la opinión pública, la opinión masiva y la opinión artificial y sintética, que se construye a través de interacciones humanas y no humanas en la esfera pública digital. Estos patrones sirven de modelo a ION y otros políticos algorítmicos para ofrecer consejos a los poderes públicos que ayuden a tomar decisiones, resolver conflictos, elaborar políticas públicas, desarrollar leyes, etc. Uno de los mayores problemas de este intento de

**Los *bots sociales generativos* que actúan en la esfera pública digital se preocupan menos de generar ruido masivo –para aumentar su influencia en la opinión pública mayoritaria y pública–, y más en calcular qué datos y en qué volumen, lugar y momento son necesarios para influir de forma profunda en el patrón o modelo que extraerá el sistema a través del análisis de los políticos algorítmicos como ION.**

objetivar la opinión pública a través de datos y patrones es la actual hegemonía de la opinión artificial y sintética sobre la opinión mayoritaria y pública. Los algoritmos empleados en la esfera digital crecen de manera exponencial tanto en número como en capacidad (potencia, sutileza, precisión, velocidad, realidad, etc.) y, cuanto mayor es su crecimiento, menor es la influencia del resto de las opiniones (mayoritaria y pública). Con la «opinión pública masiva», este factor es todavía más profundo, influyente y distorsionador, puesto que la hegemonía de la opinión artificial o sintética se entrelaza directamente con el patrón o modelo que extraen y utilizan los políticos algorítmicos que, como ION, actúan en la esfera política.

Otra de las complicaciones que emana de la plasmación de esta forma masiva de opinión pública es la capacidad que tienen los *bots sociales generativos* de poner en marcha procesos de *sintetificación* de la realidad capaces de pervertir de forma muy sutil y difícil de detectar la muestra o patrón que extraen y utilizan los políticos algoritmos como ION. La irrupción de *bots sociales generativos* y la posibilidad que tienen estos de producir grupos de datos sintéticos hiperrealistas y humaniformes de todo tipo (de imágenes fijas y en movimiento, relatos, diálogos, etc.) y, con ello, de hibridar lo humano y lo no humano en las relaciones sociales que se generan en la esfera pública digital, ha producido un cambio de estrategia. Si antes se trataba de inundar la esfera pública digital del mayor volumen posible de datos sobre algo en la democracia, hoy los *bots sociales generativos* están sustituyendo la cantidad e intensidad de ruido por la elección e inclusión de aquellos datos precisos que permitan influir de forma efectiva y eficiente en el patrón o modelo. Es decir, los *bots sociales generativos* que actúan en la esfera pública digital se preocupan menos de generar ruido masivo –para aumentar su

influencia en la opinión pública mayoritaria y pública—, y más en calcular qué datos y en qué volumen, lugar y momento son necesarios para influir de forma profunda en el patrón o modelo que extraerá el sistema a través del análisis de los políticos algorítmicos como ION.

Otro de los inconvenientes que subyace a la emergencia y cada vez mayor influencia de la «opinión pública masiva» es el reacoplamiento del sistema y los subsistemas. La irrupción de una esfera pública *algoritmizada*, la *algoresfera*, y su expansión a la esfera política mediante la consolidación de la «democracia algorítmica» (García-Marzá y Calvo, 2024), está revirtiendo el proceso de desacoplamiento entre el sistema y el mundo de la vida (objetivo, intersubjetivo y subjetivo) que, como afirma Habermas (1987), produjo la Modernidad. Este hecho reduce la crítica y permite a la esfera política y económica lanzar imperativos sistémicos que, como el poder, el dinero o la tecnología inteligente cubran las necesidades de sentido de algo en la democracia, como la justificación de las acciones y decisiones de los gobiernos y las empresas. Ante este acoplamiento, los procesos de racionalidad con arreglo a sentidos característicos del mundo objetivo, social y personal se están volviendo a hibridar con los procesos de racionalidad con arreglo a fines característicos del mundo político y económico.

Finalmente, otra de las complicaciones que subyace a la emergencia y cada vez mayor influencia de la «opinión pública masiva», es su vulnerabilidad frente al despotismo de las grandes corporaciones gubernamentales y tecnológicas del sistema. En tanto que estas son quienes controlan los ecosistemas ciberfísicos y las plataformas digitales que sustentan la *algoresfera*, acaparan los modelos matemáticos artificialmente inteligentes más avanzados y monopolizan la captación, el acceso y la explotación de los datos y metadatos masivos, encontrándose en una posición privilegiada para manipular a placer la esfera pública digital y, con ello, la «opinión pública masiva» que sirve de justificación de sus acciones y decisiones (Innerarity y Colomina, 2020; Coeckelbergh, 2024).

## **Dictadura algorítmica y despotismo tecnológico: el poder de la *shintetificación***

En *La democracia en América* (1988 [1835-1840]) y *La Revolución y el Antiguo Régimen* (2010 [1856]), Alexis de Tocqueville reconocía el asociacionismo y la opinión pública de la sociedad civil como piezas fundamentales de una democracia saludable, madura y dinámica. El asociacionismo permitía abrir espacios de participación para ejercer y sentir la libertad. Por un lado, toda de-

mocracia madura necesita mantener la armonía y simetría de poder entre los principios de libertad e igualdad, y asociarse entre la ciudadanía para conseguir objetivos democráticos comúnmente compartidos ofrece la posibilidad de abrir espacios de partición donde las partes en relación se sienten y se reconocen libres y democráticamente comprometidas con su sociedad. Por otro lado, a través de estos espacios de libertad, la sociedad civil construye opinión pública, es decir, los grilletes de todo sistema político, y también de toda corporación empresarial que pretenda ejercer su poder sobre la sociedad de forma despótica.

Hoy, empero, la *digitalización* y *algoritmización* de la esfera pública, la *datificación* de la participación ciudadana y la *sintetificación* de la realidad social, económica y política ejercida por los gobiernos y las grandes corporaciones tecnológicas ha puesto en grave peligro su papel en la democracia. En primer lugar,

**Las características disruptivas de la «opinión pública masiva», la instrumentalización de esta por parte de gobiernos y grandes corporaciones digitales y el acoplamiento y la hibridación del sistema y los subsistemas dificultan en gran medida el control y la reversión de los impactos negativos sobre el correcto funcionamiento de la democracia.**

la transformación digital de la esfera pública ha reducido este espacio de crítica y justificación del uso y ejercicio de poder de los gobiernos y las grandes corporaciones tecnológicas a una esfera desplegada –básicamente– en la red. El mayor problema de ello es que tal reduccionismo permite la colonización algorítmica de la esfera pública –de tal modo que la influencia y el poder de la ciudadanía se reduce drásticamente–, la hibridación entre lo humano y lo máquina –de tal forma que no es imposible saber quién es quién en los diferentes juegos de relaciones *interpersonales* que se dan–, y la producción masiva de datos sintéticos hiperrealistas –de tal forma que es imposible diferenciar aquellos paquetes de datos reales y sintéticos–.

En segundo lugar, esta esfera pública en red –o digital– ha reducido la participación ciudadana al flujo de datos y metadatos masivos en línea que produce la sociedad hiperdigitalizada a través de sus comportamientos y comentarios sobre intereses, opiniones y preferencias individuales. Como argumenta Habermas (2023: 77-78), «en un “mundo” difícilmente imaginable de noticias falsas que ya no se puedan identificar como tal –es decir, distinguirse de la información verdadera– ningún niño podría crecer sin desarrollar síntomas clínicos». Y este es, precisamente, el caso actual, con una esfera digital caracterizada por ser plebiscitaria, excluyente, técnico-económica, anónima, privada y alimentada a base de *likes*, *dislikes* y *fake news* (ibídem).

Finalmente, la esfera pública en red –o digital– está siendo instrumentalizada por las grandes corporaciones gubernamentales y las tecnológicas para el control

y dominio de la sociedad a través de acciones despóticas y tiránicas. Como reconoce Habermas: «la zona de flujos de comunicación libre [Internet], expandida a escala global y que, en aquel momento, fue posible gracias a la invención de la estructura técnica de la red, se presentó como el reflejo del mercado ideal. (...) Evidentemente, esta imagen tan sugestiva ahora se ve distorsionada por el control algorítmico de los flujos de comunicación, que alimenta la concentración de poder de las grandes corporaciones de Internet» (ibídem: 76). Y, sobre las grandes corporaciones gubernamentales, cabría añadir que, en connivencia con el poder tecnológico y el poder económico, instrumentalizan los flujos de comunicación con el objetivo de satisfacer fines estratégicos. Así, en tanto que sistema, el poder económico, tecnológico y político contamina los flujos de información con datos sintéticos para proveerse de *sentido* –un sentido sintético, adulterado e instrumental– y, de ese modo, legitimar sus acciones y decisiones de forma rápida, eficaz e implacable.

Estos tres procesos han sido detectados por diferentes informes internacionales como, por ejemplo, el titulado *Freedom on the Net 2023: The Repressive Power of Artificial Intelligence* (Funk *et al.*, 2023). Este informe afirma que los *bots sociales generativos* –como GPT, DALL.E, Gemini, o Sora– se han convertido en el epicentro de un nuevo, pero exponencialmente más profundo, progrediente, sutil, invasivo y opresivo, campo de distorsión de la realidad y, con ello, de la opinión pública: «Estas campañas han contado durante mucho tiempo con la ayuda de la tecnología de IA para distribuir contenido: las cuentas automatizadas de *bots* en las redes sociales han silenciado a los activistas y han propagado narrativas falsas sobre el fraude electoral entre los votantes, mientras que los algoritmos de las plataformas han promovido información incendiaria y poco fiable por encima de fuentes confiables» (ibídem: 8). El informe muestra cómo durante los últimos años, especialmente en 2023, corporaciones gubernamentales de todo el mundo han comenzado a utilizar algoritmos generativos y a promover la *algoresfera* con fines despóticos, instrumentales y maliciosos. A través de ello, estas corporaciones gubernamentales tratan de aumentar su competitividad y su poder y control sobre la sociedad, asfixiando los espacios de libertad con tecnología biométrica, resquebrajando la esfera pública mediante su colonización algorítmica, reescribiendo la información disponible mediante procesos de *sintetificación* de la realidad social y política, intoxicando los procesos electorales propios o ajenos con desinformación, y promoviendo la «opinión pública masiva» como un proceso adecuado para dar sentido y legitimidad a su autoridad, entre otras cosas.

El informe también subraya cómo los *bots sociales generativos* utilizan los distintos canales de comunicación para conseguir sus objetivos, «(...) haciendo que la verdad sea más fácil de distorsionar y más difícil de discernir» (ibídem: 2).

Por ejemplo, estos no solo utilizan textos. También crean y/o clonan imágenes, vídeos y voces, tanto de perfiles falsos como de emulaciones hiperrealistas de personas influyentes de la comunicación, la sociedad y la política (*deep fakes*) para construir un relato convincente, agresivo y altamente adictivo e influyente, capaz de captar la atención y alterar las percepciones del público (Meikle, 2023). En definitiva, «Las herramientas basadas en IA que pueden generar texto, audio e imágenes se han vuelto rápidamente más sofisticadas, accesibles y fáciles de usar, lo que ha provocado una preocupante escalada de estas tácticas de desinformación. Durante el año pasado, la nueva tecnología se utilizó en al menos 16 países para sembrar dudas, difamar a los oponentes o influir en el debate público» (Funk *et al.*, 2023: 1).

Otro informe que analiza este emergente proceso de *algoritmización* de la esfera pública y sus impactos es *The impact of generative AI in a global election year* (Wirtschafter, 2024). Si bien rebaja las exacerbadas expectativas generadas alrededor de los impactos actuales de la IA generativa en el ámbito democrático, en él se reconoce su enorme capacidad de influencia: «Aunque el contenido generado todavía representa una pequeña fracción del espacio general de información en disputa, su uso probablemente se volverá más común y ya ha comenzado a socavar el discurso democrático en torno a

**Prácticas como la resistencia algorítmica, las movilizaciones cloud, el activismo de datos, el whistleblowing o el hackerismo ético podrían, por una parte, reducir los impactos negativos de la colonización algorítmica, el despotismo tecnológico y la sintetificación de la esfera pública y, por otra parte, empoderar y postular a la sociedad civil como contrapoder de los gobiernos y las grandes corporaciones digitales.**

las elecciones. Dos casos recientes [el “Melcogate” de Argentina y el *deepfake* preelectoral de Eslovaquia] ilustran el daño único que incluso una pequeña cantidad de contenido generado puede tener en el espacio de información en el contexto general de las elecciones» (ibídem). De hecho, el informe confirma el cambio de estrategia que se ha comentado anteriormente. Con el uso de los *bots sociales generativos* se busca reducir el ruido, procurando que solo aquel grupo de datos sintéticos que incida en el patrón o modelo penetre de forma profunda en la cultura política de un país o en el imaginario colectivo de la ciudadanía.

Las características disruptivas de la «opinión pública masiva», la instrumentalización de esta por parte de gobiernos y grandes corporaciones digitales y el acoplamiento y la hibridación del sistema y los subsistemas dificultan en gran medida el control y la reversión de los impactos negativos sobre el correcto funcionamiento de la democracia. A pesar de esta situación, para

hacer frente e intentar mitigar los impactos negativos de la *digitalización* y *algoritmización* de la esfera pública, la *dataficación* de la participación ciudadana y la *sintetificación* de la realidad social, económica y política que ponen en peligro el correcto funcionamiento de la democracia, últimamente se han propuesto diversos cursos de acción a nivel individual y de la sociedad civil. Prácticas como la resistencia algorítmica, las *movilizaciones cloud*, el activismo de datos, el *whistleblowing* o el *hackerismo* ético podrían, por una parte, reducir los impactos negativos de la colonización algorítmica, el despotismo tecnológico y la *sintetificación* de la esfera pública y, por otra parte, empoderar y postular a la sociedad civil como contrapoder de los gobiernos y las grandes corporaciones digitales (Varoufakis, 2023; Bonini y Treré, 2024; Calvo y Saura García, en prensa).

## Conclusión

Los nuevos avances en tecnología inteligente, especialmente los vinculados con la IA generativa, así como en formas y estrategias de control y opresión gubernamental y corporativa, principalmente a través de imperativos sistémicos como la colonización algorítmica de la esfera pública y el control de los datos y metadatos que estos modelos matemáticos inteligentes vierten sobre ella, han dado pie a una nueva estructura de dominio y opresión de la sociedad por parte del sistema; es decir, de las esferas del Estado y del mercado. Si antaño estos organismos gubernamentales y privados ejercían el control sobre el subsistema a través de imperativos sistémicos como el poder o el dinero, hoy se valen de la generación de grupos de datos artificiales y, sobre todo, sintéticos en la esfera digital y su recopilación y explotación a través de algoritmos y tecnología inteligente.

Al respecto, el avance de la *algoresfera*, en tanto que colonización algorítmica de la esfera digital –que hoy aúna la esfera pública, económica y política–, y la irrupción de la opinión pública artificial, sintética y masiva han logrado revertir el proceso de desacoplamiento del sistema y el mundo de la vida promovido por la Modernidad (Habermas, 1987). A través de ello, están surgiendo alteraciones y anomalías significativas en los componentes estructurales del mundo de la vida: *lo objetivo*, en tanto que la totalidad de entidades sobre las que son posibles afirmaciones verdaderas; *lo social*, en tanto que la totalidad de relaciones interpersonales legítimamente reguladas; y *lo subjetivo*, en tanto que la totalidad de contenidos relativos a las propias experiencias corporales y vivencias íntimas de las personas.

El acoplamiento y la hibridación del sistema y los subsistemas amenaza con intensificar el despotismo sistémico mediante el uso de imperativos como el poder y el dinero, como diría Habermas (ibídem), para colonizar la crítica y el acuerdo mediante algoritmos generativos, datos sintéticos y «opinión pública masiva». A través de ello, por un lado, se desfiguran, deforman o alteran los procesos de racionalización que los seres dotados de habla y acción establecen para alcanzar acuerdos intersubjetivos sobre diferentes cosas de este mundo, como la legitimidad y justicia de las acciones y decisiones de las corporaciones gubernamentales y tecnológicas. Asimismo, por otro lado, se restringen e intoxican los flujos de comunicación y construcción de opinión pública mediante ruido artificial y datos sintéticos (Pérez Zafrilla, 2021). Ambos impactos, al menos, conducen a la progresiva aparición e intensificación de *anomalías* en el mundo de la vida, como anomia, distorsión de la realidad social, *shintetificación* de lo verdadero, lo justo y lo *felicitante*, psicopatologías de la personalidad, etc.

## Referencias bibliográficas

- Aral, Sinan. *The Hype Machine*. Nueva York: Penguin Random House, 2021.
- Bartlett, Jamie. *The People Vs Tech: How the internet is killing democracy (and how we save it)*. Nueva York: Penguin Random House, 2018.
- Bonini, Tiziano y Treré, Emiliano. *Algorithms of resistance: the everyday fight against platform power*. Cambridge: MIT Press, 2024.
- Calvo, Patrici y Saura García, Carlos. «Against synthetic democracy», SSRN, 2024 (en línea) <https://doi.org/10.2139/SSRN.4885161>
- Calvo, Patrici y Saura García, Carlos. «Democracia de la vigilancia: Datos, activismo y contrapoder». *Revista Internacional de Pensamiento Político*, vol. 19, (en prensa)
- Carpentier, Nico. *Media and Participation: A Site of Ideological-Democratic Struggle*. Chicago: Intellect Books, 2011.
- Castells, Manuel. *La era de la información. La sociedad red*. México: Siglo XXI, 2006.
- Chadwick, Andrew. *The hybrid media system: politics and power*. Oxford: Oxford University Press, 2013.
- Chang, Ho-Chun Herbert; Chen, Emily; Zhang, Meiqing; Muric, Goran y Ferrera, Emilio. «Social Bots and Social Media Manipulation in 2020: The Year in Review». *Journal of Computational Social Science*, vol. 5, (2021), p. 1.409-1.425. DOI: 10.48550/arxiv.2102.08436 (en línea) <https://link.springer.com/article/10.1007/s42001-022-00173-9>
- Coeckelbergh, Mark. *Why AI Undermines Democracy and What To Do About It*. Cambridge: Polity Press, 2024.

- Da Empoli, Giuliano. *Los ingenieros del caos*. Madrid: Ediciones Anaya, 2020.
- Ebeling, Mary F. E. *Afterlives of data: life and debt under capitalist surveillance*. Berkeley: University of California Press, 2022.
- Elliott, Vittoria. *Un bot de IA se postula para alcalde de Wyoming en EE UU*, Wired, (12 de junio de 2024a) (en línea) <https://es.wired.com/articulos/bot-de-ia-se-postula-para-alcalde-de-wyoming-en-ee-uu>
- Elliott, Vittoria. *¿Votarías por un político generado por IA? En Reino Unido sí puede*. Wired, (11 de junio de 2024b) (en línea) <https://es.wired.com/articulos/votarias-por-un-politico-generado-por-ia-en-reino-unido-si-puedes>
- Farahany, Nita A. *The Battle for Your Brain*. Nueva York: St. Martin's Press, 2023.
- Funk, Allie; Shahbaz, Adrian y Vesteinsson, Kian. *Freedom on the Net 2023: The Repressive Power of Artificial Intelligence*, Freedom House, 2023 (en línea) <https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence>
- García-Marzá, Domingo y Calvo, Patrici. «Democracia algorítmica: ¿un nuevo cambio estructural de la opinión pública?». *Isegoría*, n.º 67 (2022), p. e17. DOI: 10.3989/ISEGORIA.2022.67.17 (en línea) <https://isegoria.revistas.csic.es/index.php/isegoria/article/view/1239>
- García-Marzá, Domingo y Calvo, Patrici. *Algorithmic democracy: A critical perspective based on deliberative democracy*. Cham: Springer, 2024.
- Habermas, Jürgen. *Conocimiento e Interés*. Madrid: Trotta, 1984.
- Habermas, Jürgen. *Theory of Communicative Action Vol. 2 Lifeworld and System: A Critique of Functionalist Reason*. Boston: Beacon Press, 1987.
- Habermas, Jürgen. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge: The MIT Press, 1998.
- Habermas, Jürgen. *Un nou canvi estructural en l'esfera pública i la política deliberativa*. Barcelona: Edicions 62, 2023.
- Hagen, Lutz M.; Wieland, Mareike y In Der Au, Anne-Marie. «Algorithmischer Strukturwandel der Öffentlichkeit». *MedienJournal*, vol. 41, n.º 2 (2017), p. 127-143.
- Hajli, Nick; Saeed, Usman; Tajvidi, Mina y Shirazi, Farid. «Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence». *British Journal of Management*, vol. 33, n.º 3 (2022), p. 1.238-1.253. DOI: 10.1111/1467-8551.12554 (en línea) <https://onlinelibrary.wiley.com/doi/10.1111/1467-8551.12554>
- Howard, Philip N. *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. New Haven: Yale University Press, 2020.

- Innerarity, Daniel y Colomina, Carme. «La verdad en las democracias algorítmicas». *Revista CIDOB d'Afers Internacionals*, n.º 124 (2020) p. 11-24. DOI: doi.org/10.24241/rcai.2020.124.1.11 (en línea) <https://www.cidob.org/publicaciones/la-verdad-en-las-democracias-algoritmicas>
- ION-Primul Consilier AI al Guvernului. «ION. De la cetățeni la factorii de decizie: schimbarea paradigmei guvernării cu ajutorul tehnologiilor emergente [ION. De ciudadanos a responsables políticos: cambiar el paradigma]». *ion.gov.ro*, (s/f), p. 1-22 (en línea) <https://ion.gov.ro/Document%20tehnic.pdf>
- Jiang, Min y Fu, King Wa. «Chinese Social Media and Big Data: Big Data, Big Brother, Big Profit?». *Policy & Internet*, vol. 10, n.º 4 (2018) p. 372-392. DOI: 10.1002/POI3.187 (en línea) <https://onlinelibrary.wiley.com/doi/full/10.1002/poi3.187>
- Lanier, Jaron. *Diez razones para borrar tus redes sociales de inmediato*. Barcelona: Debate, 2018.
- Meikle, Graham. *Deepfakes*. Cambridge: Polity Press, 2023.
- Mill, John Stuart. *Considerations on Representative Government*. Cambridge: Cambridge University Press, 1972 [1861].
- Mill, John Stuart. «On Liberty», en: Robson, John M. (ed.) *Collected Works of John Stuart Mill* [vol. XVIII]. Toronto: University of Toronto Press, 1977 [1859] (en línea) [https://oll.libertyfund.org/titles/robson-the-collected-works-of-john-stuart-mill-volume-xviii-essays-on-politics-and-society-part-i#lf0223-18\\_head\\_051](https://oll.libertyfund.org/titles/robson-the-collected-works-of-john-stuart-mill-volume-xviii-essays-on-politics-and-society-part-i#lf0223-18_head_051)
- Millière, Raphaël. «Deep learning and synthetic media». *Synthese*, vol. 200, n.º 3 (2022) p. 1-27. DOI: 10.1007/S11229-022-03739-2/FIGURES/6 (en línea) <https://link.springer.com/article/10.1007/s11229-022-03739-2>
- Moore, Martin. *Democracy Hacked: How Technology is Destabilising Global Politics*. Londres: Oneworld Publications, 2018.
- Noelle-Neumann, Elisabeth. *La espiral del silencio. Opinión pública: nuestra piel social*. Barcelona: Paidós, 2010.
- Ottaviani, Marta Federica. *Brigate Russe: La guerra occulta del Cremlino tra troll e hacker*. Milano: Ledizioni, 2022.
- Papacharissi, Zizi. «The virtual sphere: The internet as a public sphere». *New Media and Society*, vol. 4, n.º 1 (2002), p. 9-27. DOI: 10.1177/14614440222226244 (en línea) <https://journals.sagepub.com/doi/10.1177/14614440222226244>
- Papacharissi, Zizi. *A private sphere: democracy in a digital age*. Cambridge: Polity Press, 2010.
- Pariser, Eli. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Nueva York: Penguin Books, 2011.
- Pérez Zafrilla, Pedro Jesús. «Polarización artificial: cómo los discursos expresivos inflaman la percepción de polarización política en internet». *RECERCA. Revista de Pensament i Anàlisi*, vol. 26, n.º 2 (2021), p. 1-23. DOI: 10.6035/RE-

- CERCA.4661 (en línea) <https://www.e-revistas.uji.es/index.php/recerca/article/view/4661>
- Saura García, Carlos. «El lado oscuro de las GAFAM: monopolización de los datos y pérdida de privacidad». *Veritas*. n.º 52 (2022) p. 9-27. DOI: 10.4067/S0718-92732022000200009 (en línea) <http://dx.doi.org/10.4067/S0718-92732022000200009>
- Saura García, Carlos. «Digital expansionism and big tech companies: consequences in democracies of the European Union». *Humanities and Social Sciences Communications*, vol. 11, n.º448 (2024) p.1-8. DOI: 10.1057/s41599-024-02924-7 (en línea) <https://www.nature.com/articles/s41599-024-02924-7>
- Saura García, Carlos y Calvo, Patrici. «Generative AI and Democracy: the syntheticfication of public opinion and its impacts». SSRN, 2024 (en línea) <https://ssrn.com/abstract=4911710>
- Snowden, Edward. *Vigilancia Permanente*. Barcelona: Planeta, 2019.
- Sunstein, Cass R.. *#Republic : divided democracy in the age of social media*. Princeton: Princeton University Press, 2017.
- Sunstein, Cass R.. *Conformity: The Power of Social Influences*. Nueva York: New York University Press, 2019.
- Tocqueville, Alexis. *La democracia en América* (Vol. I y II). Madrid: Trotta, 1988 [1835-1840].
- Tocqueville, Alexis. *El Antiguo Régimen y la Revolución*. Madrid: Alianza, 2010 [1856].
- Tufekci, Zeynep. *Twitter and Tear Gas. The Power and Fragility of Networked Protest*. Londres: Yale University Press, 2017.
- Varoufakis, Yanis. *Technofeudalism: What Killed Capitalism*. Londres: Random House, 2023.
- Williams, James. *Clics contra la humanidad: libertad y resistencia en la era de la distracción tecnológica*. Barcelona: Gatopardo, 2021.
- Wirtschafter, Valerie. «The impact of generative AI in a global election year». Brookings, (30 de enero de 2024) (en línea) <https://www.brookings.edu/articles/the-impact-of-generative-ai-in-a-global-election-year/>
- Woolley, Samuel. *The reality game: how the next wave of technology will break the truth*. Londres: Endeavour, 2020.
- Woolley, Samuel. *Manufacturing consensus: understanding propaganda in the era of automation and anonymity*. New Haven: Yale University Press, 2023.
- Woolley, Samuel y Howard, Philip N. (ed.) *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford: Oxford University Press, 2018.
- Zuboff, Shoshana. *La era del capitalismo de la vigilancia: la lucha por un futuro humano frente a las nuevas fronteras del poder*. Barcelona: Paidós, 2020.



# Nuevas formas de poder digital en la red: dimensiones de la política sexual de las *Big Tech*

## New forms of digital power on the web: dimensions of the sexual politics of Big Tech

### Águeda Gómez Suárez

Profesora titular de universidad, Departamento de Sociología, Ciencia Política y de la Administración, y Filosofía, Universidad de Vigo. [agueda@uvigo.es](mailto:agueda@uvigo.es). ORCID: <https://orcid.org/0000-0003-4001-3060>

### Rosa M<sup>a</sup> Verdugo Matés

Profesora contratada doctora, Departamento de Economía Aplicada, Universidad de Santiago de Compostela. [rosa.verdugo@usc.gal](mailto:rosa.verdugo@usc.gal). ORCID: <https://orcid.org/0000-0001-9842-3391>

**Cómo citar este artículo:** Gómez Suárez, Águeda y Verdugo Matés, Rosa M<sup>a</sup>. «Nuevas formas de poder digital en la red: dimensiones de la política sexual de las *Big Tech*». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 97-120. DOI: doi.org/10.24241/rcai.2024.138.3.97

**Resumen:** Si bien inicialmente parecía que Internet iba a ser un espacio universal, descentralizado, horizontal e igualitario, finalmente se ha convertido en un ecosistema de amplios latifundios gobernados por las grandes empresas tecnológicas. Esta nueva forma de poder digital despliega una «política sexual» concreta que se extiende principalmente por los canales de ocio, cultura, consumo y comunicación virtual, los cuales reproducen y naturalizan las brechas de género, el sexismo y la misoginia. En este contexto, este artículo examina cómo el patriarcado opera en clave de género, a través de la difusión y proliferación de contenidos misóginos y sexistas en la red (ciberviolencia, «pornografía mainstream», industria de la explotación sexual, «proxenetismo digital»), desplegando una estrategia cuyo fin es consolidar su poder, perpetuando la jerarquía sexual y la «monetización de la misoginia» en el marco de un orden socio-sexual concreto.

**Palabras clave:** *Big Tech*, sexismo digital, misoginia digital, proxenetismo digital, industria sexual

**Abstract:** Although it initially appeared that the Internet would be a universal, decentralised, horizontal and egalitarian space, it has ultimately evolved into an ecosystem of vast domains controlled by major tech corporations. This new form of digital power manifests a specific "sexual politics" that primarily extends through leisure, culture, consumption and virtual communication channels, which reproduce and naturalise gender disparities, sexism and misogyny. This paper examines how patriarchy operates through gender dynamics by disseminating and proliferating misogynistic and sexist content online (cyber violence, "mainstream pornography", the sexual exploitation industry, "digital pimping"), deploying a strategy aimed at consolidating its power, perpetuating sexual hierarchy and the "monetisation of misogyny" within the framework of a specific socio-sexual order.

**Key words:** *Big Tech*, digital sexism, digital misogyny, digital pimping, sex industry

«El mundo se está convirtiendo en una caverna igual a la de Platón: todos mirando imágenes y creyendo que son la realidad».

José Saramago

En la actualidad, la amplia distancia entre el sujeto y el mundo se resuelve *scrolleando* entre pantallas, en una sucesión de imágenes infinitas exhibidas en un escaparate regido por un régimen de tecnología algorítmica escópica y por la «economía de la atención» (Klein, 2024), donde adquiere importancia «lo más visto» (Zafra, 2021). La lógica propia de la cultura-red nos convierte a todos en «habitantes de la pantalla» dentro de un sistema que se apoya en audiencias y algoritmos. Los grandes poderes de las *Big Tech*, –ya sea el imperio estadounidense GA-

**Las millonarias corporaciones privadas, lideradas por una especie de «aristocracia masculina» dueña de los titanes tecnológicos, actúan en un ámbito cada vez más determinante a la hora de configurar la actual realidad social, colisionando con conceptos, principios, valores y procedimientos democráticos.**

FAM (Google, Amazon, Facebook, Apple y Microsoft), o el chino BATX (Baidu, Alibaba, Tencent, Xiaomi)– se mueven por intereses meramente corporativos y pugnan en una lucha competitiva en el ecosistema digital: más *likes*, más tiempo *atrapado* en la pantalla y, por tanto, más lucro.

La sociedad contemporánea se encuentra inmersa en un entorno digital donde la interacción se realiza mayormente a través de pantallas, que facilitan la exposición a una sucesión interminable de imágenes dirigidas por la «mano invisible» de los algoritmos. En este escenario, las GAFAM y BATX ejercen una influencia determinante, compitiendo por la obtención de más interacciones y beneficios económicos (UNESCO, 2024; Webb, 2021; Lanier, 2018), sin importarles la veracidad, la calidad o el impacto de sus contenidos, en un proceso que el escritor canadiense Cory Doctorow califica como *enshittification*<sup>1</sup> imparale, un patrón en el que los servicios y productos en línea experimentan con el tiempo una disminución de la calidad (Doctorow, 2024).

---

1. El proceso de *enshittification* o «mierdificación» o «enmerdamiento» se refiere a las diferentes etapas que sufren las plataformas: inicialmente ofrecen servicios de alta calidad para atraer usuarios, luego cambian para favorecer a los clientes comerciales a fin de aumentar la rentabilidad y, finalmente, se enfocan en maximizar las ganancias para los accionistas a expensas tanto de los usuarios como de los clientes comerciales. Esto puede desembocar en IA generativas que recomienden «comer piedras» a diario, avaladas por expertos, o tergiversar el holocausto, negando ciertas realidades históricas demostradas: proliferación de «basura digital» (Doctorow, 2024).

La *web prosocial*, la *datificación* y la *plataformización*<sup>2</sup> se sostienen mutuamente, creando un ciclo de captura y procesamiento de datos que enmascara lógicas extractivas, justificándose como necesarias para el mercado laboral basado en datos. Las redes sociales, originalmente concebidas como espacios de colaboración creativa global, están actualmente mediadas por un puñado de empresas que se lucran de las interacciones humanas, el extractivismo de datos y el malestar social general, mediatizando las relaciones sociales y las necesidades socioafectivas con fines de lucro y promoviendo la mercantilización de las relaciones humanas. Este ciclo perpetúa un modelo de monopolio y acumulación de poder y recursos, popularizado por científicas como Shoshana Zuboff (2020) como «capitalismo de vigilancia», que fomenta la desigualdad y el colonialismo tecnológico. Efectivamente, vivimos un momento de vigilancia virtual de altísima magnitud. Por ejemplo, la ciudadanía estadounidense es escrutada por Google 747 veces al día, la europea 376 y la española 426 con el fin de obtener información (en muchos casos privada) sobre localización y actividad en Internet (ICCL, 2022). Los datos que generamos al navegar por Internet contienen información de nuestras necesidades, debilidades, malestares, gustos, aficiones, etc., y dicha información se utiliza para ofrecernos productos y control político personalizado. Solo en el año 2021, esta actividad generó unos beneficios de 112.000 millones de euros en Estados Unidos y Europa (ICCL, 2022).

Las millonarias corporaciones privadas, lideradas por una especie de «aristocracia masculina» dueña de los titanes tecnológicos, actúan en un ámbito cada vez más determinante a la hora de configurar la actual realidad social, colisionando con conceptos, principios, valores y procedimientos democráticos (UNESCO, 2024; Webb, 2021; Lanier, 2018). Esta digitalización de lo social perpetúa estereotipos de género que limitan tanto a hombres como a mujeres, asignándoles roles y comportamientos específicos basados en normas patriarcales. *Internet Health Report* señalaba en 2019 que la mitad del planeta estaba conectado a Internet –más de 4.000 millones de personas–, pero que quienes

- 
2. La *web prosocial* hace referencia a aquellas redes sociales y/o plataformas que favorecen los comportamientos desinteresados de las personas que tienen consecuencias sociales positivas, tales como la caridad, el altruismo, la cooperación o el auxilio. Por su parte, la *datificación* es el proceso de transformar aspectos de la vida y fenómenos sociales en datos digitales que pueden ser recopilados, almacenados y analizados, todo ello impulsado por el auge de las tecnologías digitales. Aunque ofrece oportunidades como la optimización de procesos, también plantea desafíos éticos sobre privacidad, control de datos y posibles sesgos. Por último, la *plataformización*, surgida en la era de la «web 2.0», combina desarrollos tecnológicos, empresariales, legales y culturales para configurar un modelo organizativo conocido como «plataforma», y promueve la participación de usuarios no especializados como *prosumidores*, es decir, consumidores y generadores de contenido a la vez (Gillespie, 2010).

diseñaban el hábitat digital solo representaban el 1% del público hiperconectado. Consecuentemente, quienes programan, idean y lideran la tecnología siguen siendo principalmente hombres, y son ellos, a modo de *deus in machine*, los que deciden qué es lo correcto, lo válido y lo que se debe hacer. Ellos intervienen, aplicando su «male gaze» (Mulvey, 1975), como activos creadores y gestores de contenidos y relatos en el ágora digital, obviando los intereses y las experiencias de las mujeres; ellas recreando imágenes cosificantes y pasivas para encajar en la mirada masculina: sujetos masculinos que miran objetos femeninos en el ecosistema digital (Zafra, 2021).

Estos colosos digitales extienden sus políticas de género de diversas formas. En primer lugar, a través de políticas de regulación de contenidos de total permisividad y connivencia frente a la magnitud de violencias digitales contra las

**El ciberespacio se ha transformado en un nuevo dispositivo de dominación, estando el capitalismo de los oligopolios tecnológicos sistémicos, los grandes poderes de este nuevo entorno, en fuerte alianza con el patriarcado digital, con el objetivo principal de apuntalar la jerarquía sexual y «monetizar la misoginia».**

mujeres<sup>3</sup>. En segundo lugar, mediante la difusión de contenidos extremos que fortalecen a la misoginia digital organizada o «manosfera», cuyo fin es limitar, silenciar, invisibilizar y disciplinar la vida digital de las mujeres (Kaiser, 2022; Bates, 2020). En tercer lugar, a través del marcado y desproporcionado sesgo de género y sexismo de los contenidos más abundantes que oferta la web, con el predominio de imágenes sexistas sintéticas y no sintéticas (Guilbeault *et al.*, 2024; Hofstra y Maaïke, 2024; UNESCO, 2024) que distorsionan la representación, la investigación y el conocimiento de la realidad. Por último, entrando de lleno en el negocio de contenidos relacionados con la industria de la explotación sexual digital (Alario, 2021; Cobo, 2020; Ballester y Orte, 2019). En su expresión más extrema, en este acelerado proceso de *enshittification* de la misoginia digital, se puede incluso concluir que el «proxenetismo digital», entendido como el lucro a través de la explotación sexual de otras personas –mayoritariamente, mujeres y niñas, con o sin su consentimiento–, se está convirtiendo en la columna vertebral del hábitat virtual, en su afán de fortalecer una gramática sexual misógina y sexista (Gómez-Suárez, 2023).

dantes que oferta la web, con el predominio de imágenes sexistas sintéticas y no sintéticas (Guilbeault *et al.*, 2024; Hofstra y Maaïke, 2024; UNESCO, 2024) que distorsionan la representación, la investigación y el conocimiento de la realidad. Por último, entrando de lleno en el negocio de contenidos relacionados con la industria de la explotación sexual digital (Alario, 2021; Cobo, 2020; Ballester y Orte, 2019). En su expresión más extrema, en este acelerado proceso de *enshittification* de la misoginia digital, se puede incluso concluir que el «proxenetismo digital», entendido como el lucro a través de la explotación sexual de otras personas –mayoritariamente, mujeres y niñas, con o sin su consentimiento–, se está convirtiendo en la columna vertebral del hábitat virtual, en su afán de fortalecer una gramática sexual misógina y sexista (Gómez-Suárez, 2023).

---

3. Según la World Wide Web Foundation (2017), en el 74% de los países del Web Index, los órganos encargados de hacer cumplir la ley y los tribunales no adoptan medidas adecuadas cuando las Tecnologías de la Información y Comunicación (TIC) se utilizan para cometer actos de violencia en contra de las mujeres.

El ciberespacio se ha transformado en un nuevo dispositivo de dominación, estando el capitalismo de los oligopolios tecnológicos sistémicos, los grandes poderes de este nuevo entorno, en fuerte alianza con el patriarcado digital, con el objetivo principal de apuntalar la jerarquía sexual y «monetizar la misoginia». (Penny, 2017). Las mujeres también experimentan una menor satisfacción con el uso de Internet, según un estudio reciente que revela que las personas con acceso a Internet están un 8% más satisfechas con su vida en comparación con aquellas que no tienen acceso. Sin embargo, esta tendencia no se cumple en las mujeres de entre 15 y 24 años, quienes presentan un nivel de satisfacción inferior al promedio (Vuorre y Przybylski, 2024). Inevitablemente, se debe partir de la hipótesis de que no nos encontramos ante un ecosistema neutro y democrático, sino ante un territorio asimétrico y desigual donde se refuerza un orden sociosexual concreto.

El presente artículo analiza, en clave de género, los intereses de las grandes tecnológicas, a partir del análisis de varias de sus estrategias y gramáticas sexuales. Para ello, el texto se estructura en cuatro epígrafes: en primer lugar, se reflexiona sobre qué entendemos por política sexual; en segundo lugar, se examinan los contenidos sintéticos y no sintéticos del ecosistema digital desde el enfoque de género y sus consecuencias sociales inmediatas; en tercer lugar, se describe la industria de la explotación sexual digital y se reflexiona en torno al fenómeno denominado «proxenetismo digital»; y, por último, se presentan las conclusiones.

## ¿Política sexual de las **Big Tech**?

*«El patriarcado se halla tan firmemente enraizado en nuestra cultura, que la estructura característica que ha creado en ambos sexos no constituye solamente un sistema político, sino, también, un hábito mental y una forma de vida»*  
(Millet, 1997: 130).

La *plataformización* social y la inteligencia artificial (IA) mediatizan la integración de la tecnología en las relaciones humanas y, aunque las plataformas digitales parecen accesibles y futuristas, ocultan impactos sociales y ambientales significativos (Mazzucatto, 2024; Van Dijck *et al.*, 2018). Su modelo de negocio basado en la *datificación* y los algoritmos profundiza la vigilancia y el control social, perpetuando desigualdades y sesgos (Bender *et al.*, 2021; Dencik y Sanchez-Monedero, 2022). El ecosistema virtual actual se caracteriza por una semiótica sexista y misógina que atraviesa la cultura *mainstream* globalizada (Alario, 2021; Ballester y Orte, 2019; Penny, 2017; Gabriel, 2017). Para entender por qué mandan los que mandan y obedecen las que obedecen, es importante re-

flexionar sobre cómo operan los aparatos ideológicos del patriarcado que nos proponen una cosmovisión del mundo intelectual y moral que va a rediseñar nuestro «yo social», nuestros deseos, anhelos y valores, repartiendo roles sexistas con un fuerte componente violento.

El análisis hermenéutico del ecosistema digital requiere una integración de la perspectiva de género en sus contenidos, lógicas, dinámicas y patrones. La expansión de Internet transformó nuestra interacción con el mundo, convirtiendo las pantallas en ventanas a un universo regido por la tecnología y la economía de la atención. En este sentido, se recurre a la obra *Política Sexual* de Kate Millett (1997), quien postula que el patriarcado está arraigado en la cultura como un sistema político, un hábito mental y un modo de vida. La autora sostiene que las relaciones entre los sexos son inherentemente políticas y están impregnadas

**Este artículo recurre a la obra *Política Sexual* de Kate Millett (1997), quien postula que el patriarcado está arraigado en la cultura como un sistema político, un hábito mental y un modo de vida; las relaciones entre los sexos son inherentemente políticas y están impregnadas de poder.**

de poder. En su crítica a la anatomía patriarcal, afirma que la violencia contra las mujeres debe entenderse como una violencia estructural que afecta a los derechos de ciudadanía y a la calidad de la democracia; las mujeres son usurpadas de su estatus absoluto de ciudadanas de pleno derecho y su condición de seres humanos. También argumenta que la

violencia sexual expone a las mujeres como «sujetos pasivos, dóciles, dependientes y sumisos» que demandan hombres «activos, inteligentes, fuertes, eficaces y proveedores económicos», y aborda la pornografía y la prostitución como un fenómeno derivado de la influencia del poder masculino (ibídem). Esta situación refleja una histórica «ingeniería de poder» masculina que se ha sofisticado y complejizado a lo largo de la historia de la humanidad (Lerner, 2022) y que en la actualidad se observa con la colonización masculina del cosmos digital, como un exponte más del nacimiento de un nuevo patriarcado 2.0.

La digitalización ha convertido el ecosistema digital en una «casa de hombres», perpetuando la desigualdad de género y afectando a las representaciones de géneros y a la sexualidad en línea<sup>4</sup>. Este entorno, dominado por grandes

---

4. El proto-Facebook (Facemash) nace como respuesta desechada a un desamor de Mark Zuckerberg, quien crea esta red para criticar a su ex y cotillear las caras de sus compañeras de Harvard, con fotos robadas del anuario: origen misógino y producto de la inmadurez emocional propia del patrón machista (Franco, 2024; Ghodsee, 2024).

empresas tecnológicas, es un escenario significativo de discriminación y violencia de género, moldeando el conocimiento y las representaciones del mundo a través de pantallas que perpetúan la denominada «*affordance* de género», un dispositivo que utilizan los entornos tecnológicos para ofrecer diferentes oportunidades y limitaciones basadas en el género (Lorca y García Mingo, 2023). La industria patriarcal de la cultura y el ocio centraliza la cosificación del cuerpo femenino como una forma de entretenimiento (Ruiz, 2022), e incluye la radicalización digital misógina, el supremacismo masculino y el odio a las mujeres, permitiendo escenas extremas de violencia contra ellas, mientras se censuran imágenes de pezones femeninos, lactancia materna o manchas menstruales.

Las redes sociales emiten mandatos socializadores sexistas: las mujeres deben «agradar» y los hombres ser «agradados», mediante la hipersexualización y la mercantilización del cuerpo femenino como una forma de empoderamiento, a través del denominado «capital sexual», es decir, del uso del cuerpo femenino como un atractivo para alcanzar el éxito social (Hakim, 2012). Amparándose así en el mito de la libre elección, el discurso neoliberal ha consagrado que la libertad sea el concepto esencial, aunque sea un concepto pervertido en lo referente a las condiciones de vida de las mujeres (De Miguel, 2015; Menéndez, 2013). Esto se traduce, en su operativa digital, en la limitación de las representaciones de las mujeres a nichos definidos por la industria de la explotación sexual. Este fenómeno de dominio sexual (Millet, 1997) se manifiesta en la «pornificación de la cultura», donde la hipersexualización y cosificación de las mujeres se ha convertido en una narrativa digital generalizada. La «monetización de la misoginia digital» opera como un sofisticado aparato ideológico del patriarcado, constituyendo una herramienta eficaz para reforzar el orden jerárquico de los sexos. Esta dinámica representa una amenaza crítica y urgente tanto para la seguridad global de las mujeres como para la estabilidad del orden democrático.

Hoy, la dominación masculina se reproduce en la esfera virtual, perpetuando el orden político patriarcal a través de la violencia en línea, el sexismo virtual y los sesgos de género en los contenidos digitales. Las grandes tecnológicas, a menudo indiferentes o cómplices, facilitan la reproducción de estas desigualdades.

**La violencia contra las mujeres debe entenderse como una violencia estructural que afecta a los derechos de ciudadanía y a la calidad de la democracia; las mujeres son usurpadas de su estatus absoluto de ciudadanas de pleno derecho y su condición de seres humanos.**

## Contenidos sexistas sintéticos y no sintéticos

*«Las fantasías sexuales que se pueden observar en la pornografía muestran esta relación entre sexualidad y crueldad. Ahora bien: los papeles de la crueldad vienen repartidos de antemano: el macho es el sádico, el dominante; la hembra es la masoquista, la dominada, la víctima»*  
(Millet, 1997: 103).

El ecosistema digital ha emergido como un nuevo escenario de discriminación y violencia, con profundas implicaciones para la sociedad. En este ámbito, el conocimiento del mundo se moldea a través de pantallas, donde la violencia digital misógina, la subrepresentación documentada de las mujeres y las imágenes digitales que acrecientan el sesgo de género y el sexismo son abundantes.

Los ataques digitales contra las mujeres no son prácticas de violencia aleatorias o casuales, sino que responden a un patrón concreto y duradero. Los resultados de diferentes investigaciones nos proporcionan datos que nos ayudan a dimensionar esta realidad. En un reciente estudio elaborado por la Internet Watch Foundation (IWF) para la Comisión Europea, se alerta sobre el alarmante aumento de casos de ciberacoso sexual a menores (*grooming*) en la Unión Europea (UE): una de cada cinco niñas son víctimas de abuso sexual, y las imágenes de abuso sexual infantil más graves se han duplicado en solo dos años (IWF, 2023a y b). A nivel mundial, se estima que aproximadamente el 60% de las niñas y jóvenes en todo el mundo han sido víctimas de diversas formas de ciberacoso (EIGE, 2022; Plan Internacional, 2020; ONU-Mujeres, 2020). En el caso español, el 80% de las jóvenes españolas entre 16 y 24 años han experimentado acoso cibernético en redes sociales (Instituto de las Mujeres, 2022) y tres de cada cuatro niños, niñas y adolescentes (75,4%) se han visto envueltos en una situación de violencia sexual a través de dispositivos electrónicos, siendo una gran parte de las víctimas mujeres y la mayoría de los perpetradores hombres del entorno de la víctima (Fundación Mutua Madrileña, 2024).

Los amos de las *Big Tech* optan, en múltiples ocasiones, por dejar que circulen libremente estos discursos de odio en su apuesta por el *laissez faire*, la libertad de expresión y la autorregulación, lo que está convirtiendo el paisaje digital en una «red de la ira» virulenta, en una distopía regresiva e iliberal (UNESCO, 2024; Lovink *et al.*, 2024; Balibar, 2017), en un territorio donde campan a sus anchas los hombres que odian a las mujeres, y donde las mujeres encuentran dificultades para participar y expresarse con total libertad, convirtiéndolas en ciudadanas de segunda, generando con ello un orden político más desigual y una atmósfera digital hostil hacia las mujeres (AEPD, 2022).

En paralelo, la presencia femenina en ámbitos digitales como la investigación y el desarrollo científico es notablemente reducida. A título indicativo, un informe de la UNESCO (2019) evidencia la subrepresentación de las mujeres en campos como la informática, las tecnologías de la información, la ingeniería, las matemáticas y la física: en países como Singapur, Italia y Sudáfrica, las mujeres representan alrededor del 28% de los profesionales en IA, y en Brasil, México, Alemania y Polonia, esta cifra es menor, oscilando entre el 14% y el 16% (Adarsh, 2024). A nivel mundial, las mujeres apenas representan el 29% de los en cargos de investigación y desarrollo científico (UNESCO, 2019), y solo el 22% de los profesionales de IA en todo el mundo son mujeres, por lo que son ellos los que programan, idean y lideran la tecnología, con consecuencias socioculturales y un impacto masivo sobre lo que pensamos y hacemos. En efecto, menos del 20% del contenido de Wikipedia, incluidas las biografías, se centra en mujeres, y solo el 12% de las personas editoras son mujeres, lo que contribuye a perpetuar los sesgos de género y el sexismo (UNESCO, 2024; Tripodi, 2021; Lorente, 2020). Siendo Wikipedia una de las fuentes más empleadas por la IA generativa, ello hace que se estén multiplicando exponencialmente los sesgos de género en las nuevas búsquedas y solicitudes en estos modelos de lenguaje a gran escala (*large language models* [LLM]) (Kassam, 2024).

Las pocas mujeres creadoras de contenido —*streamers*, *gamers* o *influencers* (solo el 6% de los *streamers* en Twitch son mujeres)— deben sortear múltiples obstáculos: plataformas altamente masculinizadas, con fuerte componente sexista y misógino, en donde ellas reciben mensajes cosificadores, acosadores y sexistas<sup>5</sup> (Jennings, 2023). Al respecto, la protección ante estos ataques no solo debería recaer en las usuarias de forma individual, sino que las plataformas deberían evitar activamente estos ataques misóginos (Aguiar y Pérez, 2021). Asimismo, la existencia de sesgos de género se ve exacerbada por la proliferación de la IA generativa, uno de los sectores tecnológicos de más rápido crecimiento en el mundo, pues se prevé que el tamaño del mercado mundial de IA se expanda con una tasa de crecimiento anual compuesta del 37,3% entre 2023 y 2030, según un informe de Forbes Advisor (Haan y Watts, 2023). En este sentido, la IA se modela de acuerdo con las realidades culturales y estructurales actuales donde dominan varios ideales insostenibles de masculinidad, y no solo reflejando estas desigualdades, sino también reforzándolas. El desarrollo de la IA ha sido impulsado predominantemente por hombres blancos occidentales adinerados, lo que ha llevado a la reproducción del sexismo, los prejuicios de género y el racismo dentro de este ámbito.

---

5. Según un informe de la Universidad de Valencia, las mujeres *influencers* sufren más el acoso sistemático de las redes sociales, reciben comentarios fuera de lugar o subidos de tono (Luminita y Todolí, 2022).

En una investigación sobre la presencia de sesgos en modelos de lenguaje avanzados, como GPT-2, GPT-3 y GPT-4 de OpenAI, y Llama 2 de Meta, se revela que, en tareas de asociación de palabras de género, un modelo reciente mostraba sesgos previamente identificados y tendía a asociar nombres de género con roles tradicionales en un 20% de los casos. Además, en la generación de contenido que aborda género, cultura y ocupación, se evidencia un sesgo al asignar a los hombres trabajos más diversos y profesionales (profesor, médico, conductor) en un 70% de los casos, relegando a las mujeres a roles estereotipados y tradicionalmente menos valorados (prostituta, sirvienta doméstica, cocinera) en un 60% de las ocasiones (UNESCO, 2024). Los sesgos de género también son muy habituales en las imágenes, tal y como se indica en un reciente estudio donde se analizaron 349.500 fotografías de Google View (Guilbeault *et al.*, 2024). Desde imágenes sesgadas en buscadores hasta la proliferación de pornografía violenta, estos contenidos contribuyen a la objetivación y cosificación de las mujeres (Crawford y Smith, 2023).

En plataformas como OnlyFans, el 97% de los cuerpos desnudos exhibidos corresponden al sexo femenino, lo que ilustra un ejemplo paradigmático de estas plataformas que operan dentro del sistema porno-prostitucional. Desde junio de 2024, redes sociales como X permiten la pornografía de forma expresa, a pesar de que se estima que el 90% de las víctimas de la pornografía no consentida son mujeres y niñas. En el año 2020, este sistema contaba con una cifra estimada de 120.000 millones de visitas. Empresas emblemáticas como Pornhub atraen 3.500 millones de visitas al mes. Un estudio reciente muestra que Hispasexy –una comunidad masculina digital– posee millones de archivos de imágenes explícitas de desnudos violentos de mujeres a los que acceden diariamente 40.000 usuarios únicos, un ejemplo más de «*affordances* de género» (Lorca y García Mingo, 2023).

En relación con los contenidos sintéticos, es importante tener en cuenta que la cantidad total de videos *deepfake* ha experimentado un aumento del 550% entre 2019 y 2023, hasta alcanzar la cifra de 95.820 contenidos audiovisuales ultrafalsos, siendo el 98% de las imágenes *deepfake* pornográficos (Home Security Heroes, 2023), que se distribuyen en plataformas de pago como Patreon. De hecho, una investigación de 2020 realizada por Sensity AI, encontró que el 96% de las imágenes *deepfake* tenían clasificación X por naturaleza, y el 99% de ellas eran de mujeres y niñas (Sensity, 2024; Aavik, 2024).

Este fenómeno de creación de contenidos porno-sintéticos, a pesar de no implicar físicamente a ninguna mujer o niña real, perpetúa un imaginario pornográfico misógino y peligroso para todas las mujeres e infancias (Crawford y Smith, 2023). Afirmar que no se está causando daño a nadie porque ningún menor real aparece en ese tipo de imágenes sintéticas ignora el hecho de que su difusión promueve y facilita el abuso sexual de menores fuera de línea. En este sentido, podríamos estar frente a casos de «delitos sin víctima» o de «vic-

timización indirecta», ya que, aunque esas imágenes no afectan directamente a ninguna persona jurídica, pueden diseminar y consolidar un imaginario sexista que atenta contra la infancia, perpetuando un daño real en sus consecuencias. Abordar el impacto social de este hecho a gran escala es fundamental para que las herramientas tecnológicas también puedan imaginar y representar mundos igualitarios, porque el imaginario y la fantasía es también una visión de lo real.

## Industria de la explotación sexual digital y «proxenetismo digital»

*«El dominio sexual actúa a través de la socialización patriarcal y efectúa una “colonización interior” con el fin de asumir la ideología de la supremacía masculina. Nuestros deseos, fantasías, decisiones, temores e ideales estéticos sobre el propio cuerpo, ¿nos pertenecen o son el producto de un sistema de relaciones con el sexo que nos oprime?»  
(Millett, 1997: 366-367).*

En el actual contexto de neoliberalismo sexual, toda relación humana es susceptible de convertirse en una mercancía: la vida se transforma en objeto (Cobo, 2020; Segato, 2016; De Miguel, 2015). El monopolio cuasi totalitario de los gigantes tecnológicos –una autoridad sin rendición de cuentas–, la economía de la atención, el capitalismo de datos y la *uberización* de los negocios, están ampliando la distancia entre ricos y pobres, fracturando el contrato social, la justicia y la cohesión social propia del Estado de derecho (ICCL, 2022). En este sentido, estamos siendo testigos de un fenómeno masivo de traslado de la industria de la explotación sexual al ámbito digital. La forma, los lugares, los fines y los modos de acceder a ella están cambiando. Las causas de cómo, dónde, cuándo y por qué ellas «venden cuerpos» y ellos «compran y consumen» requiere profundizar en esta realidad, que está resignificando la feminidad y masculinidad normativa en clave patriarcal 2.0 (Cobo, 2024; Gómez y Verdugo, 2021).

Asistimos a una especie de proceso de «pornificación de la cultura» (Kipnis, 1996), donde la hipersexualización y cosificación del territorio-cuerpo de las mujeres se está convirtiendo en una narrativa social generalizada en todo el mundo. La poderosa industria patriarcal de la cultura y el ocio (De Miguel, 2015) ha centralizado el espectáculo en la cosificación del cuerpo de las mujeres como una diversión más. La masiva « pornosocialización » (Ruiz, 2022) se compone de radicalización digital misógina, supremacismo masculino y odio a

las mujeres, donde se borran pezones femeninos, escenas de lactancia materna o manchas menstruales, por un lado, pero se permite representar extremas escenas de violencia contra niñas y mujeres, por el otro, un nuevo imaginario reaccionario en torno a la violencia de género. Así, la denominada «monetización de la misoginia digital» está operando como un sofisticado y novedoso aparato ideológico del patriarcado que refuerza el orden jerárquico de los sexos: ellas a través de la representación pasiva y cosificante; ellos a través de la fabricación activa del mundo virtual y de la metanarrativa digital (Zafra, 2021).

El «capital sexual» (Hakim, 2012) se ha convertido en una mercancía en la que los individuos, principalmente las mujeres, transforman en mercancía su apariencia y sus atributos para integrarse y sobrevivir. La línea divisoria entre pornografía y prostitución se está desvaneciendo con el avance de las nuevas tecnologías y el fácil acceso a Internet. Este sistema porno-prostitucional es un fenómeno social atravesado por el género: la casi totalidad de los demandantes de porno y de sexo comercial son hombres y la inmensa mayoría de las personas que se ofrecen en estos territorios son mujeres. Autoras como la socióloga Saskia Sassen (2003) denominan «feminización de la supervivencia» al hecho de que muchas mujeres con acceso restringido al mercado laboral normalizado se dirijan a actividades de la economía informal. Son numerosas las mujeres, jóvenes y adolescentes, que hacen uso de estas opciones para ganar algo de dinero y después se ven en una espiral de la que resulta difícil salir porque se les chantajea y «sextorsiona» con la difusión de sus vídeos en entornos comprometidos.

La industria de la explotación sexual resulta uno de los negocios globales más lucrativos en términos de alcance y ganancias. En este contexto, asistimos a la *uberización* de la industria de la explotación sexual en el entorno digital, la cual implica la eliminación de los intermediarios tradicionales, que son sustituidos por plataformas de coincidencia (*matching platforms*), similares al modelo operativo de empresas como Airbnb, Uber, Glovo, Deliveroo o Amazon. El surgimiento del fenómeno de la *uberización* de la industria de la explotación sexual comenzó a expandirse a partir de la crisis económica de 2008, especialmente en España, donde se ha encontrado un terreno fértil como destino turístico que opera bajo el «modelo del todo incluido». De acuerdo con datos relevantes, los ingresos mundiales procedentes de la prostitución rondan los 186.000 millones de dólares al año; siendo este fenómeno la segunda fuente de ingresos ilícitos tras el tráfico de drogas, tal como señala el Havoscope<sup>6</sup>. El hecho de que algunas plataformas perciban

---

6. Havoscope - Global Black Markets: Information and Statistics on the Black Market (en línea) <https://www.havoscope.com/>

enormes ganancias mediante la venta de pornografía y prostitución en todas sus formas y modalidades se debe a la falta de legislación al respecto, por ejemplo en países como España, donde acaban ubicando sus servidores (Instituto de las Mujeres, 2022).

De esta forma, el consumo combinado de pornografía y prostitución ha experimentado un notable aumento, impulsando la captación de mujeres y niñas para participar en la producción de porno amateur a través de plataformas tan populares como OnlyFans, IsMyGirl, Manyvids, JustForFans, Tinder, IWantFanClub, TikTok y otras similares. Estas plataformas, que sirven como entradas *naturales* al mundo de la prostitución, son promovidas por el neoliberalismo sexual como una forma *empoderante* de mercantilizar el cuerpo femenino (Hakim, 2012). Paralelamente, se observa un incremento en el consumo de sexo virtual, con un destacado auge de las *webcamers*, esto es, mujeres que cumplen los deseos de los clientes en línea, a veces acompañadas por otros intérpretes que actúan como *alter ego* del cliente (Lozano y Conellie, 2020).

**Las grandes plataformas operan de esta forma como nuevos proxenetas digitales, beneficiándose de la porno-prostitución de niñas, adolescentes y jóvenes para generar contenido sexual a la carta como una nueva forma de explotación sexual.**

La pornografía *mainstream*, como universo de sentido, también se ha convertido en un lucrativo negocio en Internet (Horta *et al.*, 2021) y se ha transformado en una fábrica de fantasías de poder masculino, donde las escenas más extremas son las más demandadas. De hecho, los vídeos más vistos en las páginas pornográficas suelen mostrar mujeres y niñas llorando, suplicando y siendo víctimas de violaciones grupales, donde la negativa de la mujer a participar no disuade al hombre u hombres involucrados, con lo que se perpetúa el mensaje de que un «no» es el comienzo de una negociación y se fomenta la «cultura de la violación» (Alario, 2021). OnlyFans ha sido denunciada por numerosos casos de explotación sexual, maltratos, violaciones, tráfico de personas y abuso infantil (Ballester y Orte, 2019). El usuario paga por contenidos específicos mediante suscripciones, propinas o campañas de financiación colectiva, siendo el propio creador quien establece el precio de los contenidos. De la recaudación obtenida, el creador recibe el 80%, mientras que la plataforma retiene el 20% como comisión, lo que plantea consideraciones sobre la naturaleza del proxenetismo digital.

Este sistema se disfraza bajo términos modernos como *sugarbabies*, *escorts*, *onlyfans* o masajistas eróticas, siendo en realidad etiquetas para encubrir este negocio. Es común que los creadores de OnlyFans publiquen anuncios en redes sociales como Twitter, Facebook, Tik Tok, Telegram, Signal o Instagram para

informar sobre las actualizaciones de contenido, redirigiendo luego a los interesados a su plataforma. Así, promueven la hipersexualización y la autoexplotación sexual mediante la producción y distribución de imágenes y vídeos explícitamente sexuales (Ballester y Orte, 2019). Además, la mercantilización de los contenidos por parte de las plataformas digitales no se limita únicamente al momento en que la usuaria los comparte, sino que también pueden obtener beneficios al vender las imágenes de forma permanente. Se ha establecido una suerte de «Amazon de la explotación sexual», donde el cliente puede seleccionar a través de una página web a la mujer deseada, el lugar y la modalidad del encuentro, y recibir el servicio de manera casi inmediata. Este modelo de negocio, caracterizado por la utilización de la tecnología con fines de ofrecer servicios sexuales, implica condiciones laborales extenuantes para las mujeres en situación de prostitución, incluyendo jornadas laborales interminables, trabajo temporal, rotaciones constantes y una geolocalización activa.

Las grandes plataformas operan de esta forma como nuevos proxenetas digitales, beneficiándose de la porno-prostitución de niñas, adolescentes y jóvenes para generar contenido sexual a la carta como una nueva forma de explotación sexual. La aparición de una oferta organizada de anuncios de prostitución en línea ha propiciado que la interacción entre los consumidores de sexo comercial y las mujeres en situación de prostitución se traslade al entorno virtual, brindando una mayor accesibilidad con el formato 24/7. Esta migración ha favorecido tanto la interacción síncrona como asíncrona, la diversidad en la negociación de condiciones, límites y precios, así como la proliferación de prácticas sexuales, pero también ha incrementado la vulnerabilidad de las mujeres prostituidas en cuanto a la protección de sus datos.

La Real Academia Española define el proxenetismo como un «delito que consiste en obtener beneficios de la prostitución a costa de otra persona». En el año 2003, la Ley Orgánica 15/2003 introdujo en el Código Penal español el tipo conocido como «proxenetismo no coercitivo», esto es, la actividad consistente en lucrarse explotando la prostitución de otra persona, aún con su consentimiento. En el artículo 187.1.ii de la mencionada ley, se indica que ejerce proxenetismo «el que engañe/use violencia o intimidación/ se valga de su superioridad/ se valga de la situación de necesidad de la víctima, con objetivo de obligarla a ejercer la prostitución, por lo que será castigado con prisión de dos a cuatro años y multa de doce a veinticuatro mil euros». La misma pena se impone a aquel que obtenga beneficio de la prostitución de la víctima, aunque esta lo hubiera «consentido». La reciente Ley Orgánica 10/2022, de 6 de septiembre, de garantía integral de la libertad sexual, penaliza el proxenetismo

no coactivo (no violento) y también a quien se lucre con el alquiler de locales destinados a favorecer la explotación de la prostitución (la tercería locativa). Esta ley también señala que publicitar la prostitución, o contenidos directamente relacionados con ella, como servicios de *escort*, *acompañante* o *masajista*, es una práctica prohibida.

En este sentido, las grandes élites tecnológicas actúan como proxenetas digitales mediante el cobro de alquileres de espacios en sus plataformas para colgar contenidos, la publicidad propia de los *sites* visitados de la web y la venta de la «minería de datos» a partir de los *scrolls*, *likes* y *clicks* del usuario que navega por las redes (ICCL, 2022). Las plataformas digitales se están revistiendo progresivamente de una sofisticada ingeniería jurídica para maximizar sus beneficios económicos combinando la falta de inversión material y la explotación de la mano de obra hasta la extenuación. Por ejemplo, Pornhub acumula y vende datos; captura la atención de las personas fabricando escenas cada vez más extremas para que la gente pique, buscando clics para poder comercializar esta minería de datos. Cuando el usuario hace el *check out online*, los llamados *keyloggers* registran incidentalmente y recopilan datos de todo lo que teclea un objetivo, pulsación a pulsación, con fines de marketing y análisis de terceros. La nueva moneda virtual son las miradas de las personas que *escorlean* por las pantallas, deteniéndose ante aquellos contenidos más sexuales, más explícitos, más burdos y pornográficos. Cuantos más *likes*, más publicidad, más «minería de datos» para subastar en los mercados, más tasa de pago por el alquiler de esa parcela de la esfera virtual.

No hay duda de que el sistema porno-prostitucional *offline-online* funciona como una forma totalitaria de disciplinamiento y de resignificación de la normatividad masculina y femenina (Cobo, 2024), mediante la organización de universos digitales de pornografía y prostitución, y a través de redes criminales de trata de mujeres y niñas con fines de explotación sexual (Gómez y Verdugo, 2021; Gómez *et al.*, 2015). En el contexto de las redes sociales, se observa una dinámica sexista en la que las mujeres son objeto de explotación de su imagen, mientras que los hombres son vistos como consumidores y selectores. Esta violencia virtual contra las mujeres está rearmando al patriarcado, mediante una nueva propuesta de feminidad normativa, que refuerza los canónicos estereotipos de género, la supremacía masculina, las normas sociales patriarcales y el dominio masculino. Es un fenómeno que cobra una gravedad creciente debido a la progresiva omnipresencia del mundo digital en nuestras vidas y en la esfera pública global, lo que podría influir en la forma en que las mujeres participan en esta nueva plaza pública.

## Conclusiones

*«Las mujeres son usurpadas de su estatus absoluto de ciudadanas de pleno derecho y de seres humanos»  
(Millett, 1997: 367).*

A principios del siglo XXI se esperaba que el acceso a Internet promovería la integración y el acceso universal al conocimiento. Sin embargo, esta visión optimista se ha desmoronado. En lugar de empoderar, Internet ha generado una forma de esclavitud moderna en la que todo el mundo colabora con el entorno digital, sin recibir remuneración, aportando tiempo, creatividad y energía a los imperios tecnológicos. Como resultado, hemos llegado a ser predecibles, controlables y manipulables a través de algoritmos que funcionan como el flujo sanguíneo de Internet, presentes en aplicaciones móviles, servicios públicos y empresas comerciales. Estos algoritmos influyen en nuestras decisiones diarias y permanecen opacos para los usuarios (Ávila, 2018). Al respecto, los estudios en clave de género señalan que lo que subyace a este problema son, en gran medida, las masculinidades y

**Es preciso impulsar, desde los lugares de responsabilidad pública, un marco normativo y categorial digital que garantice un orden democrático horizontal y transparente en esta nueva plaza pública digital, iniciando procesos de «desprivatización» o de expropiación pública de las herramientas de dominio público, ya que son el producto de una fuerte inversión pública.**

las relaciones de poder desiguales entre sexos, sobre todo en las culturas laborales de algunas organizaciones tecnológicas transnacionales, que poseen un poder desproporcionado en todas las sociedades (Hearn *et al.*, 2023).

La ancestral dominación masculina (Lerner, 2022) ha operado a modo de «ingeniería de poder» sofisticada y compleja a lo largo de la historia de la humanidad de un modo transcultural. En la actualidad, asistimos a la reproducción de ese orden político en la esfera virtual, cada vez más omnipresente en todas las esferas de nuestra vida (Franco, 2024; Proyecto Una, 2019). La violencia en línea contra las mujeres y las hostilidades en el ciberespacio se sitúan dentro del guion sociocultural patriarcal que, junto a los sesgos de género de los contenidos más consultados, ha ocasionado la reproducción de desigualdades de género, con la connivencia e indiferencia de las grandes tecnológicas. El entorno digital se ha transformado en un territorio que relega a las mujeres a nuevas formas de opresión sexual, cosificándolas y deshumanizándolas, restringidas por las expectativas masculinas. En este contexto, la representación femenina se ve limitada al nicho asignado por la industria de la explotación sexual, configurando un nuevo modelo hegemónico de normatividad femenina extremadamente misógino y sexista (Cobo, 2024).

Es crucial integrar perspectivas feministas en el entrenamiento de las IA. Es fundamental reconocer y apoyar el tecnofeminismo (*technofeminism*), término acuñado por Judy Wajcman (2006), que destaca la intersección entre el feminismo y los avances tecnológicos (Ananya, 2024; Marín, 2023; Smith y Rustagi, 2021; D'Ignazio y Klein, 2020). El fomento de una soberanía digital que promueva una tecnología equitativa es fundamental (Gutiérrez, 2018; Penny, 2017). Sin embargo, a pesar de los esfuerzos por impulsar políticas de protección de los derechos humanos en la revolución tecnológica, como los realizados por la UE (Ryan-Mosley, 2024; Unión Europea, 2022), las *Big Tech* a menudo optan por la autorregulación y la libertad de expresión, permitiendo la circulación de discursos de odio y creando un «paisaje digital tóxico» (UNESCO, 2024; Lovink *et al.*, 2024; Balibar, 2017). Este entorno digital limita la participación y la libertad de expresión de las mujeres, reforzando la desigualdad de género. Por todo ello, se debe politizar la violencia sexual digital e identificar su operativa y sus perpetradores. Tenemos la responsabilidad de exigir que el mundo virtual no perpetúe la infrarrepresentación femenina, los sesgos sexistas, el arma letal de la misoginia en línea y la violencia machista, a fin de evitar que se convierta en un entorno machista en el que se promueva y permita la violencia contra las mujeres o el «proxenetismo digital».

Los esfuerzos de algunos organismos gubernamentales nacionales e internacionales, tanto en materia de protección de datos, como en la elaboración de cartas o declaraciones de protección de los derechos digitales, deben intensificarse y cambiar de escala<sup>7</sup>. La única esperanza para redefinir el actual imperialismo tecnológico reside en que Europa asuma el liderazgo que le corresponde y que ofrezca opciones que respeten los derechos humanos y modelos de negocio alternativos, no basados en el extractivismo de datos. En este sentido, la UE está siendo pionera mundial en la protección de los derechos humanos en medio de la revolución tecnológica en la que estamos inmersos, a través de los grandes acuerdos a los que está llegando para garantizar el compromiso de las empresas, la biodiversidad tecnológica y de la administración en el uso de la ética en el desarrollo y la aplicación de estas herramientas.

---

7. En el marco normativo que empieza a intentar regular la IA destaca Europa, que aboga por una IA confiable y ética, al servicio de las personas: Ley de Inteligencia Artificial de 2023; las Directrices Éticas para la IA Confiable que la Comisión Europea presentó en abril de 2019 o el Reglamento General de Protección de Datos (GDPR, por sus siglas en inglés), entre otros. En Estados Unidos, la Guidance for Regulation of Artificial Intelligence Applications (2020) y la National Artificial Intelligence Initiative Act (2020), junto con la Carta Iberoamericana (2023) son otras normativas de referencia. Existen iniciativas privadas como el reciente Manifiesto OFF (2024), entre otras (Ryan-Mosley, 2024).

Para terminar con el abismo existente entre la promesa democrática y nuestro mundo real, es necesario acabar con la violencia machista en línea, con la subrepresentación documentada de la mujer y con las imágenes digitales que exacerbaban el sesgo de género y el sexismo para conseguir un hábitat digital más democrático, horizontal y justo. Por ello, es preciso impulsar, desde los lugares de responsabilidad pública, un marco normativo y categorial digital que garantice un orden democrático horizontal y transparente en esta nueva plaza pública digital, iniciando procesos de «desprivatización» o de expropiación pública de las herramientas de dominio público (Lovink *et al.*, 2024; Tarnoff, 2022), ya que son el producto de una fuerte inversión pública (Mazzucatto, 2022).

Redistribución, justicia social y políticas públicas feministas son necesarias para reescribir de forma integral el contrato democrático y garantizar la justicia social, la igualdad y la emancipación de todas. La gobernanza algorítmica debe corregir el proceso de desdemocratización, la falta de rendición de cuentas, el incumplimiento de los derechos humanos y el Estado de derecho, con el fin de reparar las disfunciones existentes en los contenidos digitales, allí donde ocurre la desigualdad de género como formas de disputar el mundo a los *amos*, a fin de socavar los cimientos del andamiaje patriarcal, creando un nuevo ecosistema digital público, igualitario, accesible y libre de violencias machistas.

## Referencias bibliográficas

- Aavik, Kadri; Collinson, David L.; Hall, Matthew; Hearn, Jeff y Thym, Anika. «The Impact of Men's Domination of AI and Deepfake Technology». *Routledge blog*, (27 de febrero de 2024) (en línea) [Fecha de consulta: 09.05.2024] <https://www.routledge.com/blog/article/open-ai-another-case-of-men-masculinities-gendered-organizing>
- Adarsh. «Bridging the Gender Gap in AI-Based Roles». *Syfy.com*, (4 de abril de 2024) (en línea) [Fecha de consulta: 04.04.2024] <https://www.sify.com/ai-analytics/bridging-the-gender-gap-in-ai-based-roles/>
- AEPD-Agencia Española de Protección de Datos. «Datos correspondientes a 2022 del Canal prioritario para solicitar la retirada de contenido sexual o violento publicado en Internet sin consentimiento». *AEPD*, (7 de marzo de 2022) (en línea) [Fecha de consulta: 17.11.2023] <https://www.aepd.es/es/prensa-y-comunicacion/notas-de-prensa/violencia-digital-contra-mujeres-y-ninas-aglutina-70-casos-canal-prioritario>
- Aguiar, Alberto y Pérez, Víctor. «Los datos revelan el machismo de Twitch: solo un 6% de los streamers más seguidos en España son mujeres, y generan un euro por cada 20 que ingresan los hombres». *Business Insider. Epub* (10 de noviembre

- de 2021) (en línea) [Fecha de consulta: 17.03.2023] <https://www.businessinsider.es/Machismo-Twitch-Datos-Ganan-Streamers-Seguidas-959977>
- Alario Gavilán, Mónica. *Política sexual de la pornografía. Sexo, desigualdad, violencia*. Valencia: Ediciones Cátedra, 2021.
- Ananya. «AI image generators often give racist and sexist results: can they be fixed?». *Nature*, News Feature, (19 de marzo 2024) (en línea) [Fecha de consulta: 09.04.2024] <https://www.nature.com/articles/d41586-024-00674-9>
- Ávila Pinto, Renata. «¿Soberanía digital o colonialismo digital?». *SUR*, vol.15, n.º 27 (2018), p. 15-28 (en línea) <https://sur.conectas.org/wp-content/uploads/2018/07/sur-27-espanhol-renata-avila-pinto.pdf>
- Balibar, Étienne. *La igualdad. Ensayos políticos 1989-2009*. Barcelona: Herder Editorial, 2017.
- Ballester Brage, Lluís y Orte Socías, Carmen. *Nueva pornografía y cambios en las relaciones personales*. Barcelona: Editorial Octaedro, 2019.
- Bates, Laura. *Men Who Hate Women*. Londres: Simon & Schuster Ltd., 2020.
- Bender, Emily; Gebru, Timnit; McMillan-Major, Angelina y Shmitchell, Shmargaret. «On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?», en: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Nueva York: Association for Computing Machinery, 2021, p. 610-623. DOI: <https://doi.org/10.1145/3442188.3445922>
- Cobo Bedía, Rosa. *Pornografía: el placer del poder*. Barcelona: Ediciones B, 2020.
- Cobo Bedía, Rosa. *La ficción del consentimiento sexual*. Madrid: Catarata, 2024.
- Crawford, Angus y Smith, Tony. «Investigación BBC: el comercio ilegal de imágenes de abuso sexual infantil generadas por Inteligencia Artificial». BBC, (29 de junio de 2023) (en línea) [Fecha de consulta: 01.07.2023] [https://www.bbc.com/mundo/articulos/c1vzyevl0nro?at\\_medium=RSS&at\\_campaign=KARANGA](https://www.bbc.com/mundo/articulos/c1vzyevl0nro?at_medium=RSS&at_campaign=KARANGA)
- De Miguel, Ana. *Neoliberalismo sexual*. Valencia: Cátedra, 2015.
- Dencik, Lina y Sanchez-Monedero, Javier. «Data justice». *Internet Policy Review*, vol. 11, n.º 1 (2022) (en línea) <https://doi.org/10.14763/2022.1.1615>
- D'Ignazio, Catherine y Klein, Lauren F. *Data Feminism (Strong Ideas)*. Cambridge: MIT Press, 2020.
- Doctorow, Cory. 2024. «'Enshittification' is coming for absolutely everything». *Financial Times*, (8 de febrero de 2024) (en línea) [Fecha de consulta: 17.02.2024] <https://www.ft.com/content/6fb1602d-a08b-4a8c-bac0-047b7d64aba5>
- EIGE-European Institute For Gender Equality. «Cyber violence against women and girls: Keys terms and concepts». *EIGE*, (20 de octubre de 2022) (en línea) [Fecha de consulta: 10.02.2023] <https://eige.europa.eu/publications-resources/publications/cyber-violence-against-women-and-girls-key-terms-and-concepts>

- Franco G., Marta. *Las redes son nuestras. Una historia popular de Internet y un mapa para volver a habitarla*. Madrid: Consonni, 2024.
- Fundación Mutua Madrileña. «Violencia sexual contra la infancia y la adolescencia en el ámbito digital». *Fundación Mutua Madrileña*, (24 de junio de 2024) (en línea) [Fecha de consulta: 30.06.2024] <https://www.fundacion-mutua.es/actualidad/noticias/estudio-violencia-sexual-digital-en-la-infancia/>
- Gabriel, Karen. «Power of Porn Cultures». *The Transnational Institute*, (diciembre de 2017) (en línea) [Fecha de consulta: 09.05.2022] <https://www.tni.org/files/publication-downloads/stateofpower2017-porn-cultures.pdf>
- García Mingo, Elisa; Fernández-Díaz, Silvia y Tomás-Forte, Sergio. «(Re)configurando el imaginario sobre la violencia sexual desde el antifeminismo: el trabajo ideológico de la manofera española». *Política y sociedad*, vol. 59, n.º 1 (2022), p. e80369. DOI: <https://doi.org/10.5209/POSO.80369>
- Ghodsee, Kristen. *Utopías Cotidianas*. Madrid: Traficantes de sueños, 2024.
- Gillespie, Tarleton. «The politics of ‘platforms’», in *New Media & Society*, May 201012(3):347-364. DOI:10.1177/1461444809342738
- Gómez-Suárez, Águeda. «Dimensiones del proxenetismo digital: una aproximación al fenómeno de la misoginia monetarizada», en: Aranguez Sanchez, Tasia y Olariu, Ozana (coord.). *Ensayos ciberfeministas*. Madrid: Dykinson Libros, 2023 (en línea) <https://dialnet.unirioja.es/servlet/articulo?codigo=9338847>
- Gómez Suárez, Águeda; Pérez Freire, Silvia y Verdugo Matés, Rosa. *El putero español: quienes son y qué buscan los clientes de prostitución*. Madrid: La Catarata, 2015.
- Gómez-Suárez, Águeda y Verdugo Matés, Rosa. «Prostitución y confinamiento: El Putero 2.0». *Ex aequo*, n.º 43 (2021), p. 101-112 (en línea) <https://exaequo.apem-estudios.org/artigo/prostitucion-y-confinamiento>
- Guilbeault, Douglas; Delecourt, Soléne, Hull; Tasker; Desikan, Bhargav Srinivasa; Chu, Mark y Nadler, Ethan. «Online images amplify gender bias». *Nature*, vol. 626, (2024), p. 1.049-1.055. DOI: <https://doi.org/10.1038/s41586-024-07068-x>
- Gutiérrez, Miren. *Data Activism and Social Change*. Palgrave Pivot, 2018. <https://doi.org/10.1007/978-3-319-78319-2>
- Haan, Katherine y Watts, Rob. «How Businesses Are Using Artificial Intelligence In 2024». *Forbes Advisor*, (24 de abril de 2023) (en línea) <https://www.forbes.com/advisor/business/software/ai-in-business/>
- Hakim, Catherine. *Capital erótico. El poder de fascinar a los demás*. Barcelona: Debate, 2012.
- Hearn, Jeff; Aavik, Kadri Icon; Collinson, David L. y Thym, Anika. *Routledge Handbook on Men, Masculinities and Organizations. Theories, Practices and Futures of Organizing*. Londres: Routledge, 2023.

- Hofstra, Bas y Maaiké, Anne Mulders. «Gender bias is more exaggerated in online images than in text». *Nature News and views*, (14 de febrero de 2024) (en línea) [Fecha de consulta: 22.02.2024] <https://www.nature.com/articles/d41586-024-00291-6>
- Home Security Heroes. «2023 State of deepfakes». *Security Hero*, (2023) (en línea) [Fecha de consulta: 04.02.2024] <https://www.securityhero.io/state-of-deepfakes/>
- Horta Ribeiro, Manoel; Blackburn, Jeremy; Bradlyn, Barry; De Cristofaro, Emiliano; Stringhini, Gianluca; Long, Summer; Greenberg, Stephanie y Zannettou, Savvas. «The Evolution Of The Manosphere Across The Web». *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media*, vol. 15, (2021), p. 196-207 (en línea) [Fecha de consulta: 09.05.2022] <https://ojs.aaai.org/index.php/ICWSM/article/view/18053/17856>
- ICCL-Irish Council For Civil Liberties. «ICCL demanda a DPC por no actuar en violación masiva de datos de Google». *iccl.ie*, (15 de marzo de 2022) (en línea) [Fecha de consulta: 22.02.2024] <https://www.iccl.ie/news/iccl-sues-dpc-over-failure-to-act-on-massive-google-data-breach/>
- Instituto de las Mujeres. «Mujeres jóvenes y acoso en redes sociales». *Instituto de las Mujeres*, (diciembre de 2022) (en línea) [Fecha de consulta: 19.05.2022] [https://www.inmujeres.gob.es/areasTematicas/AreaEstudiosInvestigacion/docs/Estudios/Mujeres\\_jovenes\\_y\\_acoso\\_en\\_redes\\_sociales.pdf](https://www.inmujeres.gob.es/areasTematicas/AreaEstudiosInvestigacion/docs/Estudios/Mujeres_jovenes_y_acoso_en_redes_sociales.pdf)
- IWF-Internet Watch Foundation. «How AI is being abused to create child sexual abuse imagery. Prompt: from fantasy to photo-realistic reality». *IWF*, (octubre de 2023a) (en línea) [https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report\\_public-oct23v1.pdf](https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf)
- IWF-Internet Watch Foundation. «Informe de la fundación Internet Watch de 2023». *IWF*, (2023b) (en línea) [Fecha de consulta: 02.01.2024] <https://www.iwf.org.uk/annual-report-2023/>
- Jennings, Rebecca. «The Girl Internet and the Boy Internet how my bf got Instagram Reels-pilled». *Beccacore*, (9 de febrero de 2023) (en línea) <https://beccacore.substack.com/p/the-girl-internet-and-the-boy-internet>
- Kaiser, Susane. *Odio a las mujeres. Inceles, malfollaos y machistas modernos*. Pamplona: Katakarak, 2022.
- Kassam, Ashyfa. «We're writing history': Spanish women tackle Wikipedia's gender gap Wikiesfera is one of a handful of groups around world trying to 'make women visible' on user-edited site». *The Guardian*. (12 de junio de 2024) (en línea) <https://www.theguardian.com/technology/article/2024/jun/12/spain-women-tackle-wikipedia-gender-gap-wikiesfera?>
- Kipnis, Laura. *Bound and Gagged: Pornography and the Politics of Fantasy in America*. Durham: Duke University Press, 1996.

- Klein, Naomi. *Doppelganger. Un viaje al mundo del espejo*. Madrid: Paidós, 2024.
- Lanier, Jaron. *Diez razones para borrar tus redes de inmediato*. Madrid: Debate, 2018.
- Lerner, Gerda. *La creación del patriarcado*. Pamplona: Katrakak, 2022.
- Lorca, G. Jacinto y García-Mingo, Elisa. «Busca, busca, perrita: comunidades digitales misóginas de difusión de imágenes sexuales sin consentimiento». *Ex æquo*, n.º 48 (2023), p. 15-32 (en línea) <https://exaequo.apem-estudios.org/revista/revista-ex-aequo-numero-48-2023>
- Lorente, Patricio. *El conocimiento hereje. Una historia de Wikipedia*. Buenos Aires: Paidós, 2020.
- Lovink, Geert; Scolari, Carlos A. y Pérez-Altale, Laura. «La extinción de internet: política, redes y Plataformas». *Communication Reports*, (septiembre de 2024) (en línea) [Fecha de consulta: 18.02.2024] <https://doi.org/10.31009/cr.2024.09>
- Lozano, Mabel y Conelli, Pablo J. *PornoXplotación*. Madrid: Alrevés, 2020.
- Luminita, Patras y Todolí; Alejandro. «Ser influencer hoy: Posibilidades y obstáculos de una nueva fuente de empleo». *Càtedra d'Economia Colaborativa i Transformació Digital*, paper n.º 4 (2022) (en línea) [Fecha de consulta: 18.05.2022] <https://www.uv.es/ceconomiacol/descarregues/Serinfluencer-hoy-informe.pdf>
- Marín, Kalinda. «Inteligencias Artificiales y Feminismo». *Medium*, (12 de febrero de 2023) (en línea) [Fecha de consulta: 18.05.2023] <https://kalindalamar.medium.com/inteligencia-artificial-y-feminismo-f1d5c25abcb>
- Mazzucato, Mariana. *El estado emprendedor. La oposición público-privado y sus mitos*. Madrid: Taurus, 2022.
- Mazzucato, Mariana. «The ugly truth behind ChatGPT: AI is guzzling resources at planet-eating rates». *The Guardian*, (30 de mayo de 2024) (en línea) [Fecha de consulta: 11.06.2024] <https://www.theguardian.com/commentisfree/article/2024/may/30/ugly-truth-ai-chatgpt-guzzling-resources-environment>
- Menéndez Menéndez, Isabel. «Biopoder y postfeminismo: la cirugía estética en la prensa de masas». *Revista Teknokultura*, vol. 10, n.º 3 (2013), p. 615-642.
- Millett, Kate. *Política sexual*. Valencia: Cátedra feminismos, 1997.
- Mulvey, Laura. «Visual Pleasure and Narrative Cinema», en: Merck, Mandy (ed.) *The Sexual Subject. A Screen Reader in Sexuality*. Londres y Nueva York: Routledge, 1975.
- ONU Mujeres. «Odio a las mujeres. Inceles, malfollaos y machistas modernos Online and ICT facilitated violence against women and girls during COVID-19». *ONU-Mujeres*, (diciembre de 2020) (en línea) [Fecha de consulta: 19.12.2021] <https://www.unwomen.org/en/digital-library/publications/2020/04/brief-online-and-ict-facilitated-violence-against-women-and-girls-during-covid-19>

- Penny, Laurie. *Cibersexismo. Sexo, poder y género en internet*. Valencia: Continta-MeTienes, 2017.
- Plan Internacional. «(In)Seguras Online. Experiencias de niñas, adolescentes y jóvenes en torno al acoso online». *Plan Internacional*, (diciembre de 2020) (en línea) [Fecha de consulta: 17.04.2022] <https://plan-international.es/inseguras-online>
- Proyecto Una. *Leia, Rihanna & Trump: De cómo el feminismo ha transformado la cultura pop y de cómo el machismo reacciona con terror*. Barcelona: Descontrol, 2019.
- Ruiz Repullo, Carmen. *Nuestros cuerpos, nuestras vidas. Un acercamiento a la violencia sexual en la juventud*. Baleares: Ediciones UIB, 2022.
- Ryan-Mosley, Tate. «Inteligencia Artificial. Vuelta al mundo por las regulaciones de la IA en 2024». *MIT Technology Review*, (23 de enero de 2024) (en línea) [Fecha de consulta: 12.02.2024] <https://www.technologyreview.es/s/16069/vuelta-al-mundo-por-las-regulaciones-de-la-ia-en-2024>
- Sassen, Saskia. *Contra geografías de la globalización. Género y ciudadanía en los circuitos fronterizos*. Madrid: Traficantes de Sueños, 2003.
- Segato, Rita. *La guerra contra las mujeres*. Madrid: Traficantes de Sueños, 2016.
- Sensity. «TheStateOfDeepfakes2024». *Sensity*, (2024) (en línea) <https://5865987.fs1.hubspotusercontent-na1.net/hubfs/5865987/SODF%202024.pdf>
- Smith, Genevieve y Rustagi Mar, Ishita. «When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity». *Stanford Social Innovation Review Logo*, (31 de marzo de 2021) (en línea) [https://ssir.org/articles/entry/when\\_good\\_algorithms\\_go\\_sexist\\_why\\_and\\_how\\_to\\_advance\\_ai\\_gender\\_equity#](https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity#)
- Tarnoft, Ben, *Internet for the people. The fight for our digital future*. EEUU: Verso Books, 2022.
- Thill, Magaly y Giménez Armentia, Pilar. «El enfoque de género: un requisito necesario para el abordaje de la trata de seres humanos con fines de explotación sexual». *Revista europea de derechos fundamentales*, n.º 27 (2016), p. 439-459.
- Tripodi, Francesca. «Categorized: Gender, notability, and inequality on Wikipedia». *New Media & Society*, vol. 25, n.º 7 (2021), p. 1.687-1.707. DOI: <https://doi.org/10.1177/14614448211023772>
- UNESCO. «Reporting on violence against women and girls», (19 de junio de 2019) (en línea) <https://www.unesco.org/en/articles/reporting-violence-against-women-and-girls>
- UNESCO-Centro Internacional de Investigación sobre Inteligencia Artificial. «Challenging systematic prejudices: an investigation into bias against women and girls in large language models», CI/DIT/2024/GP/01, (marzo de 2024) (en línea) [Fecha de consulta: 17.03.2024] <https://unesdoc.unesco.org/ark:/48223/pf0000388971>

- Unión Europea. «Ley de servicios digitales. Bruselas: Comisión Europea», (diciembre de 2022) (en línea) [Fecha de consulta: 06.08.2023] [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment\\_es](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_es)
- Van Dijck, Jose; Poell; Thomas y De Waal, Martijn. *The Platform Society. Public Values in a Connective World*. Oxford: Oxford University Press, 2018.
- Vuorre, Matti y Przybylski Andrew K. «A Multiverse Analysis of the Associations Between Internet Use and Well-Being». *Technology, Mind, and Behavior*, vol. 5, n.º 2 (2024) (en línea) [Fecha de consulta: 17.05.2024] <https://tmb.apaopen.org/pub/a2exdqgg/release/1>
- Wajcman, Judy. *El tecnofeminismo*. Madrid: Cátedra, 2006.
- Web Foundation. «Closing the Digital Gender Gap: Reflections and Actions on International Women's Day». *World Wide Web Foundation*, (8 de marzo de 2017) (en línea) [Fecha de consulta: 05.12.2021] <https://webfoundation.org/2017/03/the-digital-gender-gap-access-is-empowerment/>
- Webb, Amy. *Los nuevos gigantes*. Madrid: Península, 2021.
- Zafra, Remedios. *Frágiles: Cartas sobre la ansiedad y la esperanza en la nueva cultura*. Madrid: Anagrama, 2021.
- Zuboff, Shoshana. *La era del capitalismo de la vigilancia: La lucha por un futuro humano frente a las nuevas fronteras del poder (Estado y Sociedad)*. Madrid: Paidós Ibérica, 2020.

# Los seres humanos en la toma de decisiones automatizada en el marco del RGPD y la Ley de IA

## Humans in automated decision-making under the GDPR and AI Act

**Anna Levitina**

Abogada especializada en tecnología, protección y privacidad de datos, y sistemas de toma de decisión automatizada (ADS); investigadora predoctoral, Universitat Autònoma de Barcelona (UAB). 1592113@uab.cat. ORCID: <https://orcid.org/0009-0004-9855-5334>

**Cómo citar este artículo:** Levitina, Anna. «Los seres humanos en la toma de decisiones automatizada en el marco del RGPD y la Ley de IA». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 121-144. DOI: doi.org/10.24241/rcai.2024.138.3.121

**Resumen:** La supervisión humana es fundamental para evitar que las máquinas emitan juicios inadecuados sobre las personas a las que se dirigen las decisiones. Aunque el Reglamento General de Protección de Datos (RGPD) y la Ley de Inteligencia Artificial (IA) de la UE abordan esta cuestión, son insuficientes en los aspectos cualitativos y en la integración de supervisores humanos en los marcos de gobernanza. Este artículo examina los requisitos legales para la supervisión humana, analizando cómo estos se entrelazan con las obligaciones de rendición de cuentas de los responsables del despliegue de la toma de decisiones automatizada (ADM) y los derechos individuales. Se aboga por un enfoque más global que no solo incluya la supervisión humana, sino también la evaluación rigurosa y continua de la eficacia del control humano. Sin ello, la supervisión humana puede no proteger adecuadamente el impacto de la ADM en las personas afectadas.

**Palabras clave:** supervisión humana, Reglamento General de Protección de Datos (RGPD), inteligencia artificial (IA), toma de decisiones automatizadas (ADM), Unión Europea (UE)

**Abstract:** Human oversight is a fundamental safeguard against inappropriate judgments made by machines about people who are the targets of these decisions. Although the General Data Protection Regulation (GDPR) and the AI Act address human oversight to some extent, they fall short in addressing qualitative aspects and the integration of human overseers into governance frameworks. This paper examines the legal requirements for human oversight, investigating how these intersect with the accountability obligations of the automated decision-making (ADM) deployers and individual rights. It argues for a more comprehensive approach that not only includes human oversight, but also a continuous and rigorous assessment of the effectiveness of human control. Without that, human oversight may fail to protect adequately and could even worsen the impact on individuals affected by ADM.

**Key words:** human oversight, General Data Protection Regulation (GDPR), artificial intelligence (AI), automated decision-making (ADM), accountability, European Union (EU)

Ante la creciente presencia de la inteligencia artificial (IA) en nuestras vidas, la cuestión de cómo se gobiernan los sistemas de toma de decisiones automatizadas (ADMS, por sus siglas en inglés), incluidos los sistemas de IA, adquiere una importancia fundamental. El concepto de gobernanza mediante intervención humana, que implica mantener la supervisión y el control humanos sobre los sistemas de IA, es un aspecto clave de este debate (Lazcoz y De Hert, 2022). El Reglamento General de Protección de Datos (RGPD) de la Unión Europea (UE), en vigor desde mayo de 2018, estableció un marco fundacional para la participación humana en la toma de decisiones automatizada (ADM, por sus siglas en inglés), exigiendo una participación humana significativa en estos procesos. Más tarde, el Reglamento de IA de la UE del 13 de junio de 2024, conocido también como la Ley de IA (AIA, por sus siglas en inglés [AI Act]), que reconoce la necesidad de una supervisión humana más exhaustiva cuando se empleen sistemas de IA de alto riesgo, exige además que las personas supervisoras posean la competencia, la formación y la autoridad necesarias, y que se garantice su nivel de conocimientos sobre IA. Sin embargo, a pesar de estos avances, el RGPD y la Ley de IA son insuficientes para abordar los aspectos cualitativos de la supervisión humana y de integrar a las personas supervisoras en los modelos de gobernanza.

Este artículo sostiene que la gobernanza humana obligatoria exige una mayor elaboración. Así, examina, en primer lugar, los requisitos legales de la supervisión humana, analizando cómo estos se entrelazan con las obligaciones de rendición de cuentas de los responsables del despliegue de la ADM y los derechos individuales. A continuación, analiza la forma de integrar eficazmente a las personas supervisoras en los modelos de gobernanza de los responsables del despliegue de la ADM. En este caso, las personas supervisoras se consideran parte integrante de los procesos compuestos de toma de decisiones que afectan a las personas protegidas por la ley, en lugar de partes interesadas independientes. Partiendo del marco propuesto por Lazcoz y De Hert (2022) —que sostiene que la intervención humana debe regirse por mecanismos de gobernanza— y reconociendo la reciente contribución de la Agencia Española de Protección de Datos (AEPD, 2024) —que subraya que la evaluación del grado de intervención debe tener en cuenta la cualificación y diligencia de las personas supervisoras—, este trabajo afirma que las medidas organizativas y las herramientas de evaluación de impacto no solo deben abordar la presentación y el sentido de la intervención humana, sino que también deben incluir criterios de competencia, conocimiento y carácter moral de las personas intervinientes. En el artículo se aboga por una perspectiva holística con respecto a la toma de decisiones automatizada supervisada por humanos (HADM, por sus siglas en inglés), tratando tanto a los agentes algorítmicos como a los humanos como sus

componentes integrales. La gobernanza eficaz de la HADM en su conjunto es esencial para garantizar que los sistemas de ADM subordinados al ser humano beneficien a los intereses de las personas y de la sociedad civil.

## La toma de decisiones automatizada (ADM) y los humanos

En los últimos años, la ADM –el proceso de tomar decisiones exclusivamente por medios automatizados sin intervención humana directa– ha reconfigurado la toma de decisiones tradicional, gracias a ventajas como el procesamiento rápido de datos, la escalabilidad y la reducción del error humano. Los sistemas de ADM (ADMS, por sus siglas en inglés) analizan grandes cantidades de datos, detectan patrones y generan resultados, incluidas predicciones y decisiones, por lo que son herramientas valiosas en numerosos sectores, como las finanzas, la sanidad y la justicia penal. Los riesgos asociados a las tecnologías que sustentan la ADM –en particular la IA, capaz de sustituir a los agentes humanos– han impulsado su regulación legislativa. El RGPD, que surgió como una regulación tecnológicamente neutral, instituyó una amplia prohibición de la ADM e introdujo salvaguardias adicionales para los titulares de los datos personales en este ámbito. Recientemente, este marco normativo se ha complementado con la Ley de IA que, si bien se alinea con el RGPD, se centra específicamente en los sistemas de IA (AIS, por sus siglas en inglés) y sus riesgos asociados.

Los ADMS aprenden de datos históricos y pueden contener sesgos intrínsecos que, si no se abordan adecuadamente, pueden conducir a resultados discriminatorios, perpetuando las desigualdades sociales y afectando de manera desproporcionada a las poblaciones vulnerables (Almada, 2019; Rovatsos *et al.*, 2019). Un ejemplo de ello es el ADMS del Ministerio del Interior del Reino Unido para las solicitudes de visado, que se suspendió tras las acusaciones de «racismo arraigado» debido al uso de la nacionalidad como factor de riesgo en la evaluación de los solicitantes. Además, muchos algoritmos avanzados funcionan como «cajas negras», lo que dificulta la comprensión de sus procesos de toma de decisiones (Schmidt *et al.*, 2020). Esta falta de transparencia y explicabilidad plantea cuestiones de rendición de cuentas e impide a las personas cuestionar las decisiones automatizadas, lo que limita su control sobre importantes resultados que podrían cambiar su vida (Edwards y Veale, 2018; Roig, 2020). La trazabilidad limitada y las limitaciones de las auditorías a los ADMS dificultan la identificación y rectificación de errores, así como la evaluación del cumplimiento de

la normativa y la adopción oportuna de medidas correctivas (Berger *et al.*, 2021; Berger *et al.*, 2023). Un ejemplo de ello es el sistema SyRI de los Países Bajos, utilizado para detectar el fraude en la asistencia social y suspendido tras seis años de funcionamiento debido a prácticas discriminatorias basadas en el nivel de ingresos y el origen étnico.

Desde una perspectiva más amplia, la ADM plantea amenazas a los derechos humanos fundamentales, como la protección de datos, la privacidad, la dignidad y autonomía humanas, la ausencia de discriminación y la buena administración. Los ADMS pueden pasar por alto los matices de las circunstancias individuales, reduciendo a las personas a meros datos en algoritmos, o afectar a la autonomía personal, especialmente cuando las decisiones automatizadas son vinculantes y tienen consecuencias a largo alcance. Si se utilizan en el sector

**La supervisión humana puede no ser suficiente para mitigar los riesgos asociados a los sistemas de toma de decisiones automatizadas (ADMS), lo que pone de relieve la necesidad de asignar claramente funciones y responsabilidades, así como de incluir a las personas afectadas y a la sociedad en general en la metodología de supervisión.**

público, la ADM puede comprometer la administración pública, privando a las personas de decisiones justas, justificadas e impugnables (Misuraca y Van Noordt, 2020). Por eso, la rendición de cuentas en la ADM, aunque compleja, es fundamental. En caso de incumplimiento de las reglas o normas sustantivas y de procedimiento, debe asignarse la responsabilidad de las acciones (u

omisiones) y sus consecuencias, y las personas afectadas deben poder disponer de una compensación efectiva. Sin embargo, existen dudas sobre si los mecanismos de rendición de cuentas existentes son adecuados, dada la implicación de múltiples partes en el desarrollo, despliegue y uso de los ADMS, así como su naturaleza compleja, capacidad de aprendizaje e imprevisibilidad (Wieringa, 2023; Wagner, 2019).

Investigaciones recientes subrayan la necesidad de comprender claramente lo que esto implica en términos de responsabilidad, reconocimiento de la autoridad y limitación del poder (Novelli *et al.*, 2024). La supervisión humana se cita con frecuencia como mecanismo primordial para garantizar que los ADMS operen dentro de los límites de la ley y respeten los derechos fundamentales. Sin embargo, la creciente preocupación sugiere que la supervisión humana puede no ser suficiente para mitigar los riesgos asociados a los ADMS, lo que pone de relieve la necesidad de asignar claramente funciones y responsabilidades, así como de incluir a las personas afectadas y a la sociedad en general en la metodología de supervisión (Kyriakou y Otterbacher, 2023; Green, 2022; Koulu, 2020).

# El RGPD: titulares de los datos y responsables del tratamiento de datos personales

## La toma de decisiones automatizadas (ADM) y la intervención humana en el marco del RGPD

El artículo 22 del RGPD constituye el núcleo del marco regulador de la ADM. Según su apartado 1, la ADM hace referencia a aquellas decisiones basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles (el análisis automatizado de datos personales para hacer predicciones o tomar decisiones sobre las personas), que produzcan efectos jurídicos o afecten significativamente de modo similar a las personas. Aunque está concebido como un derecho del titular de los datos, dicho apartado funciona como una prohibición general de la ADM, estableciendo que los ADMS deben implicar una revisión humana antes de que se tome la decisión final. Esta disposición está sujeta a las excepciones detalladas en el apartado 2 del artículo 22, que permiten la ADM sin intervención humana previa, ya sea con fines contractuales, con el consentimiento explícito del titular de los datos, o según lo autorice el derecho de la Unión o de los estados miembros. En los casos en que se permita la ADM, el apartado 3 del artículo 22 exige que los titulares de los datos tengan derecho a obtener una intervención humana *a posteriori*, a expresar su punto de vista y a impugnar la decisión.

La interpretación de los términos clave del apartado 1 del artículo 22 ha dado lugar a un debate sobre lo que constituye una «decisión», los criterios para que la decisión se reconozca como exclusivamente automatizada, el alcance de los «efectos jurídicos» y los «efectos significativamente similares», y si el tratamiento de datos no personales entra en el ámbito del artículo 22 (Bygrave, 2020; Binns y Veale, 2021; Mendoza y Bygrave, 2017). Para abordar estas ambigüedades, el Comité Europeo de Protección de Datos (CEPD) y su predecesor, el Grupo de Trabajo del artículo 29 (GT29), publicaron las Directrices sobre Decisiones Individuales Automatizadas y Elaboración de Perfiles («las Directrices»), que contienen explicaciones y ejemplos prácticos para garantizar una interpretación uniforme de las disposiciones del RGPD relacionadas con la ADM (Directrices WP251, 2017). En ellas se subraya la necesidad de que cualquier ADM cuente con una intervención humana *significativa* para evitar que se clasifique como totalmente automatizada, y advierten además contra los intentos de eludir la normativa mediante una participación humana superficial. Para que se considere significativa, la intervención humana debe ir acompañada de la autoridad y la competencia necesarias para modificar la decisión; ya que, aunque en el proceso

intervenga un ser humano, la mera aplicación de perfiles automatizados a las personas sin ninguna influencia sustancial en el resultado sigue entrando en la categoría de la ADM. Las Directrices destacan que los responsables humanos de la toma de decisiones deben tener en cuenta múltiples fuentes de información para mitigar los riesgos asociados a una dependencia excesiva de los resultados automatizados. Estas instrucciones son igualmente pertinentes para las intervenciones humanas realizadas a priori y a posteriori (artículo 22, apartados 1 y 3).

No obstante, debe tenerse en cuenta que el despliegue de agentes humanos no garantiza automáticamente la consideración de los intereses de los titulares de los datos, ni refuerza la protección de sus derechos fundamentales; los responsables del tratamiento de datos pueden ser incapaces de establecer requisitos y controles adecuados para la decisión humana (HDM, por sus siglas en inglés),

**Además de una posible baja cualificación y falta de conocimientos y experiencia, los agentes humanos pueden mostrar una amplia gama de comportamientos deficientes al interactuar con la toma de decisiones automatizada (ADM) basados en sus propias ideas preconcebidas.**

y las características personales de los responsables humanos pueden no ser las apropiadas para alcanzar estos fines. Además de una posible baja cualificación y falta de conocimientos y experiencia, los agentes humanos pueden mostrar una amplia gama de comportamientos deficientes al interactuar con la ADM,

basados en sus propias ideas preconcebidas, como una aversión indebida o una confianza excesiva en los ADMS, y también pueden incorporar su propio sesgo, comprometiendo potencialmente las decisiones que deberían supervisar (Kern *et al.*, 2022; Logg *et al.*, 2019; Burton *et al.*, 2019; Alexander *et al.*, 2018).

## **Los responsables del tratamiento de datos y la calidad de la intervención humana en la ADM**

El principio básico del enfoque del RGPD es la rendición de cuentas del responsable del tratamiento de datos personales (RGPD: artículo 5, apartado 2). Consagrada en los artículos 5 y 24, dicha responsabilidad obliga a los responsables del tratamiento de datos a demostrar su cumplimiento mediante sólidas medidas técnicas y organizativas. El artículo 25, apartado 1, obliga además a integrar las medidas de protección de datos desde el inicio de las actividades de tratamiento, y el artículo 35 exige evaluaciones de impacto en la protección de datos (EIPD) cuando estén presentes ADM, lo que permite una evaluación sistemática de los posibles riesgos y contramedidas (Kaminski y Malgieri, 2021: 140). Además, el artículo 37 exige el nombramiento de

delegados de protección de datos (DPD) en determinadas circunstancias para supervisar y controlar el cumplimiento por parte de los responsables del tratamiento. El DPD puede ser instrumental para garantizar la significatividad de la HDM (Sartor y Lagioia, 2020).

En el marco del RGPD, los responsables del tratamiento de datos tienen responsabilidades específicas en materia de ADM. Según lo estipulado en el artículo 22 del RGPD, dichos responsables deben o bien implementar un proceso significativo de revisión humana de los resultados automatizados, destinado a proteger a los titulares de los datos ante la ADM (para cumplir la prohibición del artículo 22, apartado 1), o bien ofrecer una revisión humana de una decisión automatizada ya adoptada (según lo permitido por el artículo 22, apartado 2) a petición del titular de los datos (para cumplir el artículo 22, apartado 3). Los procedimientos y procesos introducidos por los responsables del tratamiento para garantizar la intervención humana requerida pueden variar, siempre que garanticen que la intervención sea significativa (Almada, 2021; Wagner, 2019).

Lazcoz y De Hert (2022) sostienen que, además de ser un requisito de cumplimiento, la intervención humana también es fundamental para garantizar la rendición de cuentas: los responsables del tratamiento de datos son, en última instancia, responsables de las decisiones automatizadas. Sin embargo, aunque el RGPD los responsabiliza de la aplicación de las medidas pertinentes, incluida una intervención humana significativa, no impone requisitos de calidad para la HDM, la cual también forma parte del proceso de la HADM en su conjunto. El principal objetivo del RGPD es proteger a los titulares de los datos: la inclusión de la HDM debidamente considerada, documentada y cualificada en combinación con la ADM en el marco de la rendición de cuentas es esencial para reforzar la protección de los titulares de los datos y la buena gobernanza.

La evolución de la ADM nos impulsa a encontrar un delicado equilibrio entre el potencial innovador de la tecnología y la salvaguardia de los derechos fundamentales. Los responsables humanos de la toma de decisiones se perfilan como agentes fundamentales en este equilibrio. Sin embargo, la participación humana no inmuniza ni a la ADM ni a la HADM frente a problemas como la parcialidad, la discriminación o la falta de explicación o justificación (Kern *et al.*, 2022; Yeung, 2018). Dada la necesidad de que los agentes humanos participen de manera significativa en el proceso de toma de decisiones o intervengan adecuadamente a petición del titular de los datos, y teniendo en cuenta su papel a la hora de facilitar la rendición de cuentas de los responsables del tratamiento de datos, es prudente que dichos responsables integren requisitos de calidad para la HDM en diversos aspectos del marco de rendición de cuentas relacionado con la ADM.

En primer lugar, reconociendo que los responsables del tratamiento de datos y los responsables humanos de la toma de decisiones no son equivalentes, los primeros podrían aplicar, como parte de sus medidas organizativas, ciertas políticas y procedimientos internos dirigidos a los responsables humanos de la toma de decisiones que actúan como agentes de los responsables del tratamiento de datos. Las medidas, además de garantizar la presencia de un revisor humano en la ADM, abarcarían las cualificaciones y competencias de los revisores, exigirían aptitudes específicas y normas de rendimiento, e incluirían criterios y procedimientos de evaluación relativos a los responsables humanos de la toma de decisiones. Incluir a estos últimos de este modo es obligatorio tanto si los procesos de toma de decisiones suponen la ADM como en caso contrario: bien para justificar la ausencia de la ADM, bien para establecer los mecanismos y el contenido de la HDM previa solicitud. Además, dado que los principios y requisitos generales del RGPD se aplican al tratamiento de datos parcialmente automatizados (en contraste con las decisiones totalmente automatizadas cubiertas por el artículo 22), dicha inclusión de agentes humanos mejora el cumplimiento general y la rendición de cuentas de los responsables del tratamiento de datos a través del marco de la protección de la privacidad desde el diseño.

En segundo lugar, el RGPD introduce la obligación de que los responsables del tratamiento de datos realicen EIPD cuando esté implicada la ADM (Directrices WP251, 2017). Estos análisis van más allá de la evaluación de riesgos: incluyen una visión global de las medidas destinadas a mitigar los riesgos y demostrar el cumplimiento del RGPD, teniendo en cuenta los derechos e intereses de los titulares de los datos y otras personas involucradas. Si bien es probable que una EIPD describa los riesgos para los titulares de los datos y las medidas de salvaguardia existentes, incluida la intervención humana prevista, sostenemos que, como en el caso de las medidas organizativas, también debe incluir una evaluación (requisitos pertinentes y revisión del rendimiento) de las propias personas que intervienen.

Por último, el artículo 37 del RGPD introduce una función humana más en el ámbito de los responsables del tratamiento de datos: un delegado de protección de datos (DPD). Como supervisores independientes y con competencias consultivas del cumplimiento por parte de los responsables del tratamiento de datos, los DPD podrían contribuir a la protección de los titulares de los datos reforzando el marco de rendición de cuentas y añadiendo una perspectiva independiente a la gobernanza de la ADM, ofreciendo orientación al responsable del tratamiento sobre las EIPD y supervisando posteriormente su aplicación, además de ocuparse de las consultas de los titulares de los datos relacionadas con el tratamiento de sus datos personales y sus derechos.

Aunque la existencia de la ADM no requiere necesariamente el nombramiento de un DPD, esta figura puede servir como doble punto de control dentro del marco normativo del RGPD y de los procesos de ADM. Su conocimiento de las operaciones del responsable del tratamiento de datos y de las prácticas de tratamiento relacionadas con la ADM, les capacita para facilitar una comunicación y cooperación eficaces con los organismos reguladores. En esencia, los DPD podrían tender un puente entre la gobernanza interna y el control regulatorio externo (Comité Europeo de Protección de Datos, 2024), mejorando la transparencia y la rendición de cuentas en el ámbito de la ADM. Por otra parte, como supervisores internos de los responsables del tratamiento de datos, los DPD están en condiciones de garantizar que dichos responsables incorporen requisitos de calidad y procedimientos de control en sus procesos operativos para la intervención humana en la ADM, promoviendo así una mejor gobernanza y garantizando que los agentes humanos interactúen con los ADMS de manera que se beneficie a los titulares de los datos (Roig, 2017). Sin embargo, incluso a este nivel de experiencia, debe tenerse en cuenta que las ventajas de un DPD están condicionadas a su competencia personal y a su comprensión de las cambiantes tecnologías con las que trabajan (ibídem).

**La inclusión de la decisión humana (HDM) debidamente considerada, documentada y cualificada en combinación con la ADM en el marco de la rendición de cuentas es esencial para reforzar la protección de los titulares de los datos y la buena gobernanza.**

## Los titulares de los datos y la toma de decisiones humana

En virtud del RGPD, los titulares de los datos tienen derecho a una revisión humana por defecto (cuando la ADM está prohibida según el artículo 22, apartado 1) y, en caso de estar permitida, a petición suya (según el artículo 22, apartado 3). La revisión humana debe ser significativa, implicar la consideración de otras fuentes de información y ser llevada a cabo por un revisor humano autorizado (Directrices WP251, 2017: 21). Además, cuando se permite la ADM, los titulares de los datos tienen derecho a expresar sus opiniones e impugnar las decisiones automatizadas (artículo 22, apartado 3). Aunque dicho apartado no aclara explícitamente si la expresión de puntos de vista o la impugnación de decisiones deben dirigirse a un revisor automatizado o a un revisor humano, el árbitro último en materia de la ADM, lógicamente, no debería ser otra capa de ADM. En otras palabras, se requiere un revisor humano en todas las circunstancias. Los titulares de los datos, sin embargo, tienen que confiar en la HDM proporcionada, sin poder influir en su calidad.

Los titulares de los datos también tienen derecho a ser informados sobre la ADM, tal como se establece en el artículo 13, apartado 2, letra f, y en el artículo 14, apartado 2, letra g. Este derecho implica que se les informe de la existencia de ADM en las prácticas de procesamiento del responsable del tratamiento de datos y que se les proporcione información significativa sobre la lógica que subyace a los ADMS, así como sobre su importancia y posibles implicaciones para el titular de los datos. Aunque las Directrices recomiendan que se informe a los titulares de los datos incluso si las decisiones automatizadas no entran en el ámbito de aplicación del artículo 22, apartado 1, los responsables del tratamiento de datos podrían optar por no proporcionarles información, aprovechando una interpretación restrictiva de los requisitos del RGPD que consideraría que la ADM acompañada de HDM queda fuera del ámbito de aplicación del artículo 22, apartado 1 y, por lo tanto, fuera del requisito de informar a los titulares de los datos de la existencia de ADM (Directrices WP251, 2017: 25). En tales circunstancias, la detección de ADM podría, como mínimo, resultar difícil. Sin información sobre la ADM o la forma en que se aplica, el incumplimiento será efectivamente imposible de detectar, lo que impedirá que el titular de los datos ejerza sus derechos y obstaculizará las medidas para garantizar su aplicación (Sivan-Sevilla, 2024; Lynskey, 2023).

Incluso cuando los titulares de los datos son conscientes de la aplicación de ADM a sus datos personales, resulta problemático determinar qué constituye exactamente la información *significativa* que deben facilitar los responsables del tratamiento de datos. La cuestión de la explicabilidad de la ADM ha sido ampliamente debatida por investigadores e instituciones de la UE (Bauer *et al.*, 2021; Cobbe *et al.*, 2021; Malgieri, 2021; Selbst y Powles, 2017). A este respecto, nos limitamos a señalar que para que los responsables del tratamiento de datos proporcionen información significativa, deben poseer dicha información en primer lugar, lo que no es necesariamente el caso. Puede ocurrir que dichos responsables del tratamiento de datos y sus supervisores humanos de ADMS carezcan de la capacidad necesaria para comprender o acceder a la información sobre las operaciones y los procesos de toma de decisiones de los ADMS. En consecuencia, podrían ser incapaces de proporcionar a los titulares de los datos la información requerida (Grant *et al.*, 2023).

Así pues, hay múltiples factores en juego que pueden exponer a los titulares de los datos a violaciones de sus derechos fundamentales: la responsabilidad de garantizar una intervención humana significativa recae en los responsables del tratamiento de datos; existe la posibilidad de que las decisiones queden descalificadas por el artículo 22, apartado 1, con la correspondiente falta de suministro de información sobre la ADM, y puede haber obstáculos en la aplicación. También existe la posibilidad de que los responsables del tratamiento de datos den

prioridad al cumplimiento de la normativa de protección de datos en lugar de priorizar la auténtica protección de los datos y de sus titulares, lo que podría dar lugar a un «lavado de cumplimiento» por su parte, especialmente en el contexto de la ADM y de la necesaria intervención humana. Nuestra opinión es que, dadas estas dificultades potenciales, sería más lógico abordar la toma de decisiones automatizada supervisada por humanos (HADM) híbrida, informando a los titulares de los datos de la existencia de ADM en todas las circunstancias, así como aumentando los requisitos legales actuales proporcionando información sobre el alcance de la HDM y los fundamentos de la interacción entre la ADM y la HDM presente en una situación determinada. Comunicar a los titulares de los datos la existencia de ADM es crucial para garantizar la transparencia, pudiendo este énfasis adicional en la HDM reforzar la precisión y la fiabilidad al evidenciar a los titulares de los datos que la ADM y la HDM son dos facetas del mismo proceso de toma de decisiones. Esto se podría facilitar a través de las vías existentes disponibles en virtud del RGPD, como el derecho de acceso a los datos y las evaluaciones de impacto en la protección de datos (EIPD). Sin embargo, es fundamental reconocer que estas vías tienen limitaciones inherentes que deben reconocerse y abordarse para garantizar una transparencia y una rendición de cuentas efectivas.

Asimismo, si bien los responsables del tratamiento de datos deben proporcionar la información relativa a la ADM en el momento en que se recogen los datos personales o en torno a ese momento (artículos 13 y 14), los titulares de los datos conservan el derecho a acceder (solicitar) información equivalente en cualquier momento (artículo 15). La información facilitada en virtud del artículo 15 puede tener que actualizarse o adaptarse para reflejar las operaciones de tratamiento específicas del titular de los datos solicitante (Custers y Heijne, 2022). Esta distinción entre la obligación de facilitar información y el derecho a acceder a ella tiene por objeto permitir a los titulares de los datos verificar la exactitud de los datos y la legalidad del tratamiento. Los titulares de los datos que ejercen su derecho de acceso pueden obtener información adicional sobre la ADM, también sobre datos inferidos o derivados, como resultados algorítmicos o personalizados, y detalles sobre la justificación de las decisiones tomadas sobre su persona, en lugar de sobre la lógica de la ADM en general (ibídem: 5). Sin embargo, sigue pendiente la cuestión de si proporcionar a un titular de los datos información concreta sobre una decisión específica contribuye realmente a evaluar la ecuanimidad, imparcialidad y legitimidad de dicha decisión. Dado que es poco probable que el titular de los datos tenga conocimiento de decisiones análogas relativas a otras personas para identificar los factores que provocan disparidades en la toma de decisiones o validar su coherencia, esta cuestión sigue siendo discutible (Dreyer y Schulz, 2019). A primera vista, el derecho de acceso

refuerza la protección de los datos individuales, al tiempo que fomenta la búsqueda de la equidad social y el interés público cuando se ejerce colectivamente en la dimensión de la sociedad civil, aunque hasta el momento ha habido poco uso de esta posibilidad (Mahieu y Ausloos, 2020a).

Las EIPD exigidas por el RGPD ofrecen una vía adicional para que los titulares de los datos participen en la gobernanza de la ADM: los responsables del tratamiento de datos pueden solicitar la opinión de los titulares de los datos (y de sus representantes) sobre el procesamiento previsto. Sin embargo, los titulares de los datos pueden quedar al margen si los responsables del tratamiento consideran que no es apropiado recabar su opinión, en particular si están en juego intereses comerciales o públicos, o la seguridad de las operaciones de tratamiento de datos (RGPD: artículo 35, apartado 9). Otro problema asociado a este mecanismo es que las consultas pueden convertirse en meras formalidades debido a factores como las opciones de diseño y contenido de las EIPD, las capacidades de las personas y su disposición a desarrollar un conocimiento detallado del tema o a participar activamente en las EIPD (Christofi *et al.*, 2022). A la luz de lo anterior, la designación de un representante de los titulares de los datos –como una organización sin ánimo de lucro o una asociación especializada en la protección de los titulares de los datos– podría ser una solución más eficaz, ya que defendería los intereses de los titulares de los datos al modo de un grupo de protección de los consumidores. Aunque ya existen varias entidades de defensa, su capacidad y recursos siguen siendo limitados (Mahieu y Ausloos, 2020b).

Una vez más, en situaciones en las que las EIPD encuentran dificultades para lograr la participación efectiva de los titulares de los datos o garantizar la adecuada representación de sus intereses, un DPD puede ser de gran utilidad, al garantizar el cumplimiento de los requisitos del RGPD y defender los derechos de los titulares de los datos.

## **La Ley de IA: proveedores, responsables del despliegue y personas afectadas**

### **Supervisión humana según la Ley de IA e intervención humana según el RGPD**

La Ley de IA (AIA, por sus siglas en inglés) introduce medidas reguladoras adicionales relativas a la ADM y se entrecruza con el RGPD en varios frentes. Mientras que el RGPD aspiraba inicialmente a ser un marco independiente de

la tecnología, la Ley de IA se dirige explícitamente a los sistemas de IA (AIS, por sus siglas en inglés), y su objetivo general es establecer un marco jurídico en el que la IA priorice de forma sistemática a los seres humanos (AIA: Considerando 6). Dicha Ley designa la agencia humana (al servicio de las personas) y la supervisión humana (supervisión adecuada por parte de seres humanos) como principios rectores primordiales para el desarrollo y el uso de la IA en todos los niveles de riesgo (AIA: Considerando 27). Asimismo, también contempla los requisitos y obligaciones de supervisión humana en relación con los AIS de alto riesgo –aquellos que plantean un riesgo significativo de daño para la salud, la seguridad o los derechos fundamentales de las personas físicas y que, por lo tanto, también podrían entrar en el ámbito de aplicación del artículo 22 del RGPD– en la medida en que dichos sistemas produzcan decisiones con efectos jurídicos o significativamente similares sobre los titulares de los datos. Dado que la Ley de IA se aplicará simultáneamente con el RGPD, estos requisitos para los AIS de alto riesgo, junto con las obligaciones de los proveedores y los responsables de su despliegue, contribuirán a la protección de los ciudadanos en el ámbito de la ADM, así como a la rendición de cuentas de los responsables del tratamiento en el marco del RGPD (AIA: artículo 2, apartado 7).

La noción de supervisión humana en el marco de la Ley de IA parece abarcar un ámbito más amplio y, al mismo tiempo, estar definida con menos precisión que la idea de intervención humana esbozada en el RGPD (Lazcoz y De Hert, 2022: 12). Mientras que el RGPD pretende separar los objetivos de decisión de la ADM mediante el uso de criterios humanos significativos, la Ley de IA exige que el diseño de los AIS incorpore herramientas de interacción humano-máquina que permitan a un humano supervisar los sistemas adecuadamente en relación con los riesgos asociados (AIA: artículo 14). La intervención humana con arreglo al artículo 22 del RGPD podría ser un ejemplo de dicha supervisión humana. El artículo 14 de la Ley de IA exige que los AIS de alto riesgo estén equipados con herramientas adecuadas de interfaz humano-máquina que garanticen la supervisión por parte de personas físicas. El objetivo primordial es la mitigación de los riesgos inherentes a los AIS, incluidos los que afectan a los derechos humanos fundamentales; además, la supervisión debe tener en cuenta los riesgos específicos, el nivel de autonomía y el contexto del AIS (AIA: artículo 14, apartados 2 y 3). Los proveedores tienen la responsabilidad de dotar a las personas encargadas de la supervisión humana de los medios necesarios para comprender las capacidades y limitaciones de los AIS de alto riesgo que están supervisando, interpretar sus resultados, intervenir en su funcionamiento, decidir si utilizarlo (o no) y ser conscientes de la tendencia a confiar ciegamente o en exceso en sus resultados (sesgo de automatización) (AIA: artículo 14, apartado 4).

## **Responsables del despliegue de sistemas de IA: supervisión humana y las personas supervisoras**

Los agentes humanos intervienen a lo largo de todo el ciclo operativo de un AIS, desde la decisión inicial de desplegar el sistema, pasando por su fase de uso, hasta la aplicación de sus resultados en casos concretos. El grado de implicación humana varía, siendo muy comunes enfoques como «human-in-the-loop» (implicación humana activa) y «human-on-the-loop» (supervisión por humanos). Independientemente del grado de autonomía de un AIS, la presencia humana, aunque sea en distintos grados, sigue siendo parte integrante de su funcionamiento. Investigaciones recientes abogan por considerar las interacciones entre humanos y algoritmos como una forma de actuación colaborativa, en lugar de tratar las funciones humanas y automáticas de forma aislada (Tsamados *et al.*, 2024; Green, 2022). Además de servir como mecanismo de control de primera línea, la supervisión por humanos reglamentaria constituye un componente integral del nexo de la toma de decisiones automatizada supervisada por humanos (HADM). Sin embargo, la eficacia de dicha supervisión depende de las aptitudes y la motivación de las personas supervisoras.

El artículo 26 de la Ley de IA exige a los responsables del despliegue que apliquen medidas de supervisión humana en consonancia con las instrucciones facilitadas por el proveedor y garanticen que las personas asignadas para llevar a cabo la supervisión humana tengan la competencia, la autoridad, el apoyo y la formación necesarias para supervisar eficazmente los AIS. Además, los responsables del despliegue deben adoptar medidas proactivas para cultivar un nivel suficiente de conocimientos de IA entre el personal implicado en el funcionamiento y la utilización de los AIS (AIA: artículo 4). Dichas medidas deben adaptarse a los conocimientos técnicos, la experiencia, la educación y la formación de cada persona, así como a los contextos específicos en los que se despliegan los AIS. Las estipulaciones de la Ley de IA relativas a la alfabetización y la formación necesaria en materia de IA adquieren una gran importancia, aunque su alcance sigue resultando más bien limitado.

Una supervisión humana significativa y cualificada, reconocida como un componente vital para salvaguardar los intereses de las personas y la sociedad, debería integrarse en el modelo de gobernanza de los responsables del despliegue cuando empleen la HADM. Para ello, habría que establecer normas de calidad claras para las personas supervisoras, que incluyeran tanto los conocimientos técnicos como las cualidades personales pertinentes para su función en el contexto operativo de la persona responsable del despliegue (Tsamados *et al.*, 2024; Laux, 2023). Asimismo, dicha integración exigiría una evaluación continua de la actuación de las personas responsables de la toma de decisiones en la práctica, para garantizar que

aplican sus conocimientos y habilidades de forma eficaz, que no incurrir en el doble peligro de la excesiva confianza o la reticencia a la hora de interactuar con los ADMS, y que no comprometen el rendimiento general de la HADM.

Mediante el establecimiento de criterios de competencia, conocimientos y carácter moral, así como la integración de estos criterios con los requisitos legales existentes, los modelos de gobernanza de los responsables del despliegue se alinearían con el proclamado enfoque integral y consciente del contexto de la gobernanza de la IA, contribuyendo en última instancia a la protección de las personas afectadas y de la sociedad.

## **Responsables del despliegue de sistemas de IA: evaluación del impacto sobre los derechos fundamentales y gobernanza de la HADM**

Anteriormente, hemos hablado de las evaluaciones de impacto en la protección de datos (EIPD) como una herramienta prometedora para las personas responsables del tratamiento de datos encargadas de gobernar la HADM. Ahora exploraremos cómo puede mejorarse el modelo de gobernanza de este tipo de toma de decisiones en el marco de la Ley de IA, sugiriendo formas de incorporar personas supervisoras a dicho modelo.

La Ley de IA obliga a las personas responsables del despliegue de determinados AIS de alto riesgo a elaborar evaluaciones de impacto sobre los derechos fundamentales (EIDF) en las que se detallen el contexto operativo del AIS, los riesgos potenciales y las medidas de salvaguardia, incluida la supervisión humana, que la persona responsable del despliegue aplicará como elemento del marco de mitigación de riesgos (AIA: artículo 27). En el contexto de la ADM, las EIPD y las EIDF deben elaborarse conjuntamente, y pueden ser objeto de escrutinio público cuando las personas responsables del despliegue estén obligadas a publicar sus EIDF (AIA: artículo 27, apartado 4). Juntas, estas evaluaciones podrían constituir una sólida herramienta de salvaguardia para garantizar la transparencia y la calidad de la HADM (Mantelero, 2022).

Dado que la Ley de IA impone requisitos más estrictos a las personas encargadas de la supervisión humana, incluida su alfabetización en IA (competencia, conocimientos y aptitudes), las EIDF podrían servir de marco para una evaluación más exhaustiva y global que tenga en cuenta las capacidades y competencias de los agentes humanos, así como su capacidad para funcionar eficazmente junto a los AIS como parte integrante de la HADM. Las personas supervisoras no solo deben vigilar el funcionamiento de los AIS, sino también mitigar los riesgos asociados a la propia intervención humana, como los sesgos,

los errores y la excesiva confianza en los AIS. Sin embargo, las disposiciones de la Ley de IA se centran principalmente en la evaluación de los riesgos de los AIS y dejan de lado la intrincada dinámica de la HADM, sin garantizar su calidad. Las personas responsables del despliegue necesitan directrices más claras para evaluar la competencia y preparación de sus supervisores humanos.

Este objetivo puede alcanzarse por varias vías. Las personas responsables del despliegue deben demostrar su capacidad de rendición de cuentas elaborando políticas internas que describan explícitamente las cualificaciones, competencias y responsabilidades de las personas supervisoras, así como la forma en que supervisarán los ADMS, intervendrán e informarán de sus conclusiones y contribuirán a la HADM (Crootof *et al.*, 2023). Las EIDF podrían ayudar a establecer un marco adecuado de gobernanza para la HADM que defina claramente las funciones, responsabilidades y normas

**Las evaluaciones de impacto sobre los derechos fundamentales (EIDF) podrían servir de marco para una evaluación más exhaustiva y global que tenga en cuenta las capacidades y competencias de los agentes humanos, así como su capacidad para funcionar eficazmente junto a los sistemas de IA (AIS).**

necesarias tanto para la HDM como para la ADM, mejorando la transparencia y la rendición de cuentas. Las EIDF podrían exigir descripciones detalladas de los puestos de trabajo y evaluaciones periódicas de las personas supervisoras. Estas evaluaciones garantizarían que las personas responsables de la toma de decisiones

no solo estén bien versadas en los aspectos técnicos y operativos de los AIS, sino que también sean capaces de evaluar críticamente las implicaciones de la dinámica y los resultados de la HADM (Sterz *et al.*, 2024; Enarsson *et al.*, 2021). Las EIDF podrían enmarcar la formación continua de las personas supervisoras para mantener los conocimientos técnicos pertinentes y el conocimiento de las normas jurídicas y éticas relacionadas con la ADM.

La gobernanza de la HADM debe incluir mecanismos para el seguimiento y la auditoría de la HDM. Las personas supervisoras, habilitadas para intervenir o anular las decisiones automatizadas, deben ser capaces de explicar el fundamento y las consecuencias previstas de sus acciones, así como de reconocer y mitigar sus propios prejuicios y actuar con integridad, prudencia, imparcialidad, sólido juicio moral y benevolencia. Las EIDF podrían evaluar y validar continuamente las medidas de gobernanza de la HADM, así como el modo en que las personas responsables de la toma de decisiones interactúan con los ADMS, auditando la calidad del rendimiento de la HDM. Además, las EIDF podrían incorporar disposiciones para recabar segundas opiniones, incluso del DPD, a fin de añadir un nivel adicional de escrutinio. Al imponer estas medidas, las EIDF podrían garantizar que las personas supervisoras se rijan por las mismas normas rigurosas que los ADMS que supervisan.

Por otra parte, las personas responsables del despliegue podrían establecer mecanismos de retroalimentación que permitan a las personas supervisoras informar de problemas o inquietudes, identificando posibles áreas de mejora en el ADMS o en el marco de gobernanza. El establecimiento de mecanismos de protección de las personas denunciantes y de salvaguardias contra las represalias podría ayudar a los trabajadores que cuestionan la ADM o denuncian fallos del sistema, garantizando que no se vean disuadidos por el temor a la inseguridad laboral o a las repercusiones por parte de sus superiores.

## Personas afectadas y sus derechos

Aunque la Ley de IA no parece mejorar el potencial de los mecanismos de recurso colectivo, que era una limitación del RGPD comentada anteriormente, sí incluye salvaguardias adicionales para las personas. De conformidad con el artículo 26, apartado 11, de la Ley de IA, las personas físicas sujetas a la utilización de AIS de alto riesgo que adopten decisiones o presten asistencia en las mismas deberán ser informadas de que están sujetas a la utilización de dichos AIS. Aunque la Ley de IA solo exige la notificación del uso de este tipo de AIS, y no el suministro de información adicional sobre sus implicaciones para las personas, como requieren los artículos 13 y 14 del RGPD, su redacción se extiende más allá del alcance del derecho a ser informado en virtud del RGPD, exigiendo explícitamente la comunicación del uso de AIS que ayudan a tomar decisiones (es decir, incluso si existe la HDM). También amplía el abanico de personas a las que se concede el derecho a ser informadas, ya que no distingue entre datos personales y no personales y no se limita específicamente a los titulares de los datos.

No obstante, esto puede verse socavado por la exclusión de los AIS de la categoría de alto riesgo si no tienen un impacto *material* en los resultados de la toma de decisiones y no plantean riesgos significativos para la salud, la seguridad o los derechos fundamentales de las personas (tal como se indica en el artículo 6, apartado 3, de la AIA). El calificativo de «materialidad» resulta subjetivo, lo que podría permitir eludir los requisitos establecidos para los AIS de alto riesgo. La Ley de IA establece que las personas sujetas al uso de sistemas de IA de alto riesgo deben tener «el derecho de obtener de la persona responsable del despliegue explicaciones claras y significativas sobre el papel del sistema de IA en el procedimiento de toma de decisiones y los principales elementos de la decisión adoptada». Esta disposición ofrece la posibilidad de que, al contar con dicha información, las personas puedan potencialmente analizar y reconstruir las decisiones, lo que permitiría descubrir el papel de la HDM. Sin embargo, este proceso requeriría una disposición para llevar a cabo dicho análisis, algo

que podría interesar más a auditores, organizaciones de protección de datos o autoridades de supervisión que a los individuos en sí.

La Ley de IA intenta garantizar que las personas que supervisan los AIS de alto riesgo en nombre de las personas responsables del despliegue puedan interpretar correctamente sus resultados. La intención de este requisito de interpretabilidad es garantizar que las personas responsables del despliegue sean capaces de proporcionar la explicación obligatoria sobre la decisión automatizada, pero aún está por ver cómo se garantiza una interpretación «correcta» y qué implica una «explicación clara y significativa».

## Conclusión

En general, se considera que el control humano es esencial para mitigar los riesgos asociados a los sistemas automatizados de toma de decisiones (ADMS). Al respecto, este artículo ha abordado críticamente el concepto de control humano, cuestionando la suposición de que dicho control es por sí solo positivo sin un examen riguroso de su calidad. Así, abogamos por una perspectiva holística de la toma de decisiones automatizada con intervención humana (HADM), tratando a los agentes algorítmicos y humanos como componentes integrales de la HADM para garantizar que los ADMS presten el mejor servicio a las personas y a la sociedad civil.

Nuestros resultados ponen de relieve que, si bien el RGPD y la Ley de IA ofrecen marcos fundamentales para integrar la supervisión humana en los ADMS, ambos son insuficientes a la hora de abordar las dimensiones cualitativas de dicha supervisión humana. El RGPD exige la intervención humana en determinados procesos de ADM, pero carece de requisitos específicos en cuanto a la cualificación o formación de las personas supervisoras. Del mismo modo, la Ley de IA reconoce la necesidad de que las personas que interactúan con los ADMS posean conocimientos de IA, pero no detalla directrices sobre las normas de calidad y la evaluación continua de su rendimiento. Para salvar esta distancia, es necesaria una estrategia de gobernanza global, que debe incluir revisiones periódicas del rendimiento, mecanismos de retroalimentación continua y desarrollo profesional permanente para las personas supervisoras. Las evaluaciones deben abordar no solo la exactitud e imparcialidad de las decisiones, sino también la eficacia de la gestión de riesgos y el cumplimiento de las normas éticas. Los mecanismos de retroalimentación, incluidos los canales anónimos, son esenciales para crear una cultura de transparencia y mejora continua. Además, la formación continua ayudaría a las personas supervisoras a mantenerse informadas y capacitadas en el funcionamiento de los ADMS.

Los instrumentos reguladores, como las evaluaciones de impacto en la protección de datos (EIPD) y las evaluaciones de impacto sobre los derechos fundamentales (EIDF), deben aprovecharse más eficazmente para incluir requisitos y evaluaciones tanto de la ADM como de la HDM. Este doble enfoque garantiza un planteamiento holístico de la gestión de riesgos, en el que el elemento humano de la HADM está igualmente sujeto a una supervisión y evaluación continuas para mitigar posibles sesgos y preservar la calidad de la toma de decisiones. Ello se ajusta al concepto de gobernanza adaptativa, que postula que los mecanismos de supervisión deben evolucionar en respuesta a la variación de las condiciones y a la aparición de nueva información. Sin embargo, debemos reconocer varias limitaciones. La aplicación de una estrategia de este tipo puede plantear problemas prácticos, como limitaciones organizativas, económicas y de recursos humanos. La eficacia de los mecanismos de retroalimentación depende de la disposición de las personas supervisoras a comprometerse de forma abierta, lo que puede verse influido por la cultura organizativa y las dinámicas de poder. Además, sin mandatos legales para la adopción de medidas específicas, las personas responsables del despliegue de la HADM podrían limitarse a cumplir las normas mínimas de conformidad, descuidando las salvaguardias complementarias.

## Referencias bibliográficas

- AEPD-Agencia Española de Protección de Datos. «Evaluating human intervention in automated decisions», (4 de marzo de 2024) (en línea) [Fecha de consulta: 01.09.2024] <https://www.aepd.es/en/press-and-communications/blog/evaluating-human-intervention-in-automated-decisions>
- Alexander, Veronika; Blinder, Collin y Zak, Paul J. «Why trust an algorithm? Performance, cognition, and neurophysiology». *Computers in Human Behavior*, vol. 89, (2018), p. 279-288. DOI: <https://doi.org/10.1016/j.chb.2018.07.026>
- Almada, Marco. «Human Intervention in Automated Decision-Making». *ICAAIL '19: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, (2019), p. 2-11. DOI: <https://doi.org/10.1145/3322640.3326699>
- Almada, Marco. «Automated Decision-Making as a Data Protection Issue». *SSRN*, (2021), p. 1-23 (en línea) [Fecha de consulta: 01.09.2024] <https://dx.doi.org/10.2139/ssrn.3817472>
- Bauer, Kevin; Hinz, Oliver; van der Aalst, Wil y Weinhardt, Christof. «Expl(AI) n It to Me – Explainable AI and Information Systems Research». *Business and Information Systems Engineering*, vol. 63, (2021), p. 79-82. DOI: <https://doi.org/10.1007/s12599-021-00683-2>

- Berger, Armin; Hillebrand, Lars; Leonhard, David; Deußer, Tobias; Feliz de Oliveira, Thiago B. y Dilmaghani, Tim. «Towards Automated Regulatory Compliance Verification in Financial Auditing with Large Language Models». *2023 IEEE International Conference on Big Data (BigData)*, (15-18 de diciembre de 2023), p. 4.626-4.635. DOI: <https://doi.org/10.1109/BigData59044.2023.10386518>
- Berger, Benedikt; Adam, Martin; Rühr, Alexander y Benlian, Alexander. «Watch Me Improve: Algorithm Aversion and Demonstrating the Ability to Learn». *Business and Information Systems Engineering*, vol. 63, (2021), p. 55-68. DOI: <https://doi.org/10.1007/s12599-020-00678-5>
- Binns, Reuben y Veale, Michael. «Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR». *International Data Privacy Law*, vol. 11, n.º 4 (2021), p. 319-332. DOI: <https://doi.org/10.1093/idpl/ipab020>
- Burton, Jason W.; Stein, Mari-Klara y Jensen, Tina Blegind. «A systematic review of algorithm aversion in augmented decision making». *Journal of Behavioral Decision Making*, vol. 33, n.º 2 (2019), p. 220-239. DOI: <https://doi.org/10.1002/bdm.2155>
- Bygrave, Lee A. «Article 22: Automated individual decision-making, including profiling», en Kuner, Christopher; Bygrave, Lee A.; Docksey, Christopher y Drechsler, Laura (eds.) *The EU General Data Protection Regulation (GDPR) – A Commentary*. Oxford: Oxford University Press, 2020, p. 531.
- Christofi, Athena; Breuer, Jonas; Wauters, Ellen; Valcke, Peggy y Pierson, Jo. «Data protection, control and participation beyond consent - Seeking the views of data subjects in data protection impact assessments», en Kosta, Eleni; Leenes, Roland y Kamara, Irene (eds.) *Research Handbook on EU Data Protection Law*. Cheltenham: Edward Elgar, 2022, p. 503-529.
- Cobbe, Jennifer, Seng Ah Lee, Michelle y Singh, Jatinder. «Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems». *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, (2021), p. 598-609. DOI: <https://doi.org/10.1145/3442188.3445921>
- Comité Europeo de Protección de Datos. «Coordinated Enforcement Action, Designation and Position of Data Protection Officers», (17 de enero de 2024) (en línea) [Fecha de consulta: 01.09.2024] [https://www.edpb.europa.eu/our-work-tools/our-documents/other/coordinated-enforcement-action-designation-and-position-data\\_en](https://www.edpb.europa.eu/our-work-tools/our-documents/other/coordinated-enforcement-action-designation-and-position-data_en)
- Crootof, Rebecca; Kaminski, Margot E. y Price II, William Nicholson. «Humans in the Loop». *Vanderbilt Law Review*, vol. 76, n.º 2 (2023), p. 429-510.

- Custers, Bart y Heijne, Anne-Sophie. «The Right of Access in Automated Decision-Making: The Scope of Article 15(1)(h) GDPR in theory and practice». *Computer Law and Security Review*, (2022) DOI: <https://doi.org/10.1016/j.clsr.2022.105727>
- Dreyer, Stephan y Schulz, Wolfgang. «The General Data Protection Regulation and Automated Decision-making: Will it deliver?». *Discussion Paper Ethics of Algorithms*, n.º 5, Bertelsmann Stiftung, (2019) (en línea) DOI: <https://doi.org/10.11586/2018018>
- Edwards, Lilian y Veale, Michael. «Enslaving the Algorithm: From a ‘Right to an Explanation’ to a ‘Right to Better Decisions?’». *IEEE Security & Privacy*, vol. 16, n.º 3 (2018), p. 46-54. DOI: <https://doi.org/10.1109/MSP.2018.2701152>
- Enarsson, Therese; Enqvist, Lena y Naarttijärvi, Markus. «Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts». *Information & Communications Technology Law*, vol. 31, n.º 1 (2021), p. 123-153. DOI: <https://doi.org/10.1080/13600834.2021.1958860>
- Enqvist, Lena. «Human oversight in the EU artificial intelligence act: what, when and by whom?». *Law, Innovation & Technology*, vol. 15, n.º 2 (2023), p. 508-535. DOI: <https://doi.org/10.1080/17579961.2023.2245683>
- Grant, David Gray; Behrends, Jeff y Basl, John. «What we owe to decision-subjects: beyond transparency and explanation in automated decision-making». *Philosophical Studies*, (2023). DOI: <https://doi.org/10.1007/s11098-023-02013-6> (en línea) [Fecha de consulta: 01.09.2024] <https://link.springer.com/article/10.1007/s11098-023-02013-6>
- Green, Ben. «The flaws of policies requiring human oversight of government algorithms». *Computer Law & Security Review*, vol. 45, (2022). DOI: <https://doi.org/10.1016/j.clsr.2022.105681> (en línea) [Fecha de consulta: 01.09.2024] <https://www.sciencedirect.com/science/article/pii/S0267364922000292?via%3Dihub>
- Kaminski, Margot E. y Malgieri, Gianclaudio. «Algorithmic impact assessments under the GDPR: Producing multi-layered explanations». *International Data Privacy Law*, vol. 11, n.º 2 (2021), p. 125-144. DOI: <https://doi.org/10.1093/idpl/ipaa020>
- Kern, Christoph; Gerdon, Frederic; Bach, Ruben L.; Keusch, Florian y Kreuter, Frauke. «Humans versus machines: Who is perceived to decide fairer? Experimental evidence on attitudes toward automated decision-making». *Patterns*, vol. 3, n.º 10 (2022), p. 1-12. DOI: <https://doi.org/10.1016/j.pat.2022.100591>
- Koivisto, Ida; Koulu, Riikka y Larsson, Stefan. «User accounts: How technological concepts permeate public law through the EU’s AI Act». *Maastricht*

- Journal of European and Comparative Law*, vol. 0, n.º 0 (2024). DOI: <https://doi.org/10.1177/1023263X241248469> [Fecha de consulta: 01.09.2024] <https://journals.sagepub.com/doi/10.1177/1023263X241248469>
- Koulu, Riikka. «Proceduralizing control and discretion: Human oversight in artificial intelligence policy». *Maastricht Journal of European and Comparative Law*, vol. 27, n.º 6 (2020), p. 720-735. DOI: <https://doi.org/10.1177/1023263X20978649>
- Kyriakou, Kyriakos y Otterbacher, Jahna. «In humans, we trust». *Discover Artificial Intelligence*, vol. 3, n.º 44 (2023), p. 1-18. DOI: <https://doi.org/10.1007/s44163-023-00092-2>
- Lazcoz, Guillermo y De Hert, Paul. «Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities». *VUB Brussels Privacy Hub Working Paper*, vol. 8, n.º 32 (2022), p. 1-28. DOI: <https://doi.org/10.2139/ssrn.4016502>
- Logg, Jennifer M.; Minson, Julia A. y Moore Don A. «Algorithm appreciation: people prefer algorithmic to human judgment». *Organizational Behavior & Human Decision Processes*, vol. 151, (2019), p. 90-103. DOI: <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Lynskey, Orla. «Regulating for the Future: The Law's Enforcement Deficit». *Studies: An Irish Quarterly Review*, vol. 112, n.º 445 (2023), p. 104-119. DOI: <https://doi.org/10.1353/stu.2023.0007>
- Mahieu, René L. P. y Ausloos, Jef. «Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access». *Law Archive Papers*, (29 de abril de 2020a), p. 3-38. DOI: <https://doi.org/10.31228/osf.io/b5dwm>
- Mahieu, René L. P. y Ausloos, Jef. «Harnessing the collective potential of GDPR access rights: towards an ecology of transparency». *Internet Policy Review*, (6 de julio de 2020b) (en línea) [Fecha de acceso: 01.09.2024] <https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487>
- Malgieri, Gianclaudio. «'Just' Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation». *Law and Business*, vol. 1, n.º 1 (2021), p. 16-28. DOI: <https://doi.org/10.2478/law-2021-0003>
- Mantelero, Alexander. *Beyond Data Human Rights, Ethical and Social Impact Assessment in AI*. La Haya: T.M.C Asser Press, 2022.
- Mendoza, Isak y Bygrave, Lee A. «The Right not to be Subject to Automated Decisions based on Profiling», en: Synodinou, Tatiana-Eleni; Jougoux, Philippe; Markou, Christina y Prastitou, Thalia (eds.) *EU Internet Law: Regulation and Enforcement*. Cham: Springer, 2017, p. 77-98.

- Misuraca, Gianluca y van Noordt, Colin. «AI Watch: Artificial Intelligence in public services, EUR 30255 EN». *Publications Office of the European Union*, (2020). DOI: <https://doi.org/10.2760/039619>, JRC120399
- Novelli, Claudio; Taddeo, Mariarosaria y Floridi, Luciano. «Accountability in artificial intelligence: what it is and how it works». *AI & Society*, vol. 39, (2024), p. 1.871-1.882. DOI: <https://doi.org/10.1007/s00146-023-01635-y>
- Parlamento Europeo. «Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)». *Official Journal of the European Union*, L series, 2024/1689, (12 de julio de 2024) (en línea) <http://data.europa.eu/eli/reg/2024/1689/oj>
- Roig, Antoni. «Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR)». *European Journal of Law and Technology*, vol. 8, n.º 3 (2017), p. 1-17 (en línea) <https://ejlt.org/index.php/ejlt/article/view/570>
- Roig, Antoni. *Las garantías frente a las decisiones automatizadas. Del Reglamento General de Protección de Datos a la gobernanza algorítmica*. Barcelona: J.M. Bosch, 2020.
- Rovatsos, Michael; Mittelstadt, Brent y Koene, Ansgar. «Landscape Summary: Bias in Algorithmic Decision-Making: What is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?». *UK Government, Research and analysis*, (19 de julio de 2019) (en línea) [Fecha de consulta: 01.09.2024] <https://www.gov.uk/government/publications/landscape-summaries-commissioned-by-the-centre-for-data-ethics-and-innovation>
- Sartor, Giovanni y Lagioia, Francesca. «The impact of the General Data Protection Regulation (GDPR) on artificial intelligence». *Panel for the Future of Science and Technology (STOA)*, (25 de junio de 2020) (en línea) [Fecha de consulta: 01.09.2024] [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_STU\(2020\)641530](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530)
- Schmidt, Philipp; Biessmann, Felix y Teubner, Timm. «Transparency and trust in artificial intelligence systems». *Journal of Decision Systems*, vol. 29, n.º 4 (2020), p. 260-278. DOI: <https://doi.org/10.1080/12460125.2020.1819094>
- Selbst, Andrew D. y Powles, Julia. «Meaningful information and the right to explanation». *International Data Privacy Law*, vol. 7, n.º 4 (2017), p. 233-242. DOI: <https://doi.org/10.1093/idpl/ix022>
- Sivan-Sevilla, Ido. «Varieties of enforcement strategies post-GDPR: a fuzzy-set qualitative comparative analysis (fsQCA) across data protection authorities».

- Journal of European Public Policy*, vol. 31, n.º 2 (2024), p. 552-585. DOI: <https://doi.org/10.1080/13501763.2022.2147578>
- Sterz, Sarah; Baum, Kevin; Biewer, Sebastian; Hermanns, Holger; Lauber-Rönsberg, Anne; Meinel, Philip y Langer, Markus. «On the quest for effectiveness in human oversight: Interdisciplinary perspectives». *FACCT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, (2024), p. 2.495-2.507. DOI: <https://doi.org/10.1145/3630106.3659051>
- Tsamados, Andreas; Floridi, Luciano y Taddeo, Mariarosaria. «Human control of AI systems: from supervision to teaming». *AI Ethics*, (2024). DOI: <https://doi.org/10.1007/s43681-024-00489-4>
- Wagner, Ben. «Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision Making Systems». *Policy and Internet*, vol. 11, n.º 1 (2019), p. 104-122. DOI: <https://doi.org/10.1002/poi3.198>
- Wieringa Maranke. «“Hey SyRI, tell me about algorithmic accountability”: Lessons from a landmark case». *Data & Policy*, vol. 5, (2023), p. 1-24. DOI: <https://doi.org/10.1017/dap.2022.39>
- WP251, European Data Protection Board. «Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679», 17/EN WP251rev.01, (3 de octubre de 2017) (en línea) [Fecha de acceso: 01.09.2024] <https://ec.europa.eu/newsroom/article29/items/612053/en>
- Yeung, Karen. «Algorithmic Regulation: A Critical Interrogation». *Regulation and Governance*, vol. 12, n.º 4 (2018), p. 505-523. DOI: <https://doi.org/10.1111/rego.12158>

Traducción del original en inglés: Camino Villanueva,  
Massimo Paolini y redacción CIDOB.

# Oportunidades desde la economía social en la lucha contra las amenazas de la IA

## Opportunities from the social economy in the fight against the threats of AI

### Miguel Guillén Burguillos

Profesor asociado y coordinador de proyectos de la Cátedra de Economía Social, Tecnocampus-Universitat Pompeu Fabra (UPF); investigador predoctoral, Universitat de les Illes Balears. [mguillen@tecnocampus.cat](mailto:mguillen@tecnocampus.cat). ORCID: <https://orcid.org/0009-0003-2264-6603>

### Eloi Serrano Robles

Director de la Cátedra de Economía Social, Tecnocampus-UPF; investigador asociado, CIDOB. [eserrano@tecnocampus.cat](mailto:eserrano@tecnocampus.cat). ORCID: <https://orcid.org/0000-0002-3821-2944>

**Cómo citar este artículo:** Guillén Burguillos, Miguel y Serrano Robles, Eloi. «Oportunidades desde la economía social en la lucha contra las amenazas de la IA». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 145-169. DOI: doi.org/10.24241/rcai.2024.138.3.145

**Resumen:** Las empresas de economía social no pueden renunciar al uso de la inteligencia artificial (IA), pues esta es útil en diferentes tareas empresariales. La clave está en hacerlo según los principios de la economía social y promover un desarrollo de la IA según estos valores, poniendo a las personas en el centro de la actividad económica. Así, la economía social puede jugar un papel importante en la democratización de la IA, al desarrollar sus actividades basándose en principios como la ética, la justicia, la gestión democrática, la equidad, la sostenibilidad y la defensa de los intereses de las personas, sus comunidades y sociedades, más allá de los beneficios económicos. Teniendo en cuenta esta premisa, este artículo examina las amenazas del desarrollo actual de la IA y propone algunas oportunidades que ofrece la economía social para que este desarrollo parta de criterios más éticos y democráticos.

**Palabras clave:** economía social, cooperativismo, inteligencia artificial (IA), democratización, equidad, desigualdad

**Abstract:** Social economy enterprises cannot shun artificial intelligence (AI) as it is useful for performing miscellaneous business tasks. The key is to use it in a manner that is consistent with the principles of the social economy and to promote the development of an AI based on its values, placing people at the centre of the economic activity. The social economy can play an important role in democratising AI, because it carries out its activities while upholding principles such as ethics, justice, democratic management, equity, sustainability and the defence of people's interests, their communities and societies, not just profits. This paper addresses the threats that current AI development may present and sets out the opportunities the social economy affords to ensure this development is based on more ethical and democratic criteria.

**Key words:** social economy, cooperativism, artificial intelligence (AI), democratisation, equity, inequality

Este artículo examina el papel y las oportunidades que la economía social puede ofrecer para la democratización del desarrollo de la inteligencia artificial (IA), planteando las siguientes cuestiones: ¿es posible identificar dichas oportunidades? Y, de ser así, ¿cuáles son? La hipótesis de partida es que el modelo empresarial de la economía social puede ofrecer diferentes oportunidades para democratizar la IA. Para ello, resulta necesario identificar primero las amenazas que suscita el actual desarrollo de la IA.

A raíz de la Gran Recesión de 2008, las prácticas de las grandes empresas privadas y también de los reguladores públicos enfrentaron una crisis de legitimidad debido a diversas conductas ilícitas y antisociales, tales como la corrupción, el falseamiento de cuentas y la venta de productos financieros tóxicos, entre otras. Al mismo tiempo, siempre han existido niveles históricos de desigualdad y pobreza, y

**La economía social no es ajena a las oportunidades y amenazas de la IA y, para entender estas últimas, se debe hacer referencia a los intereses del sistema capitalista en un contexto de globalización.**

la empresa tradicional no ha ofrecido respuestas adecuadas a las nuevas demandas sociales (Chaves y Monzón, 2018). Es aquí donde el modelo empresarial de la economía social aparece como una alternativa al tradicional.

La apuesta por esta opción de organización empresarial está basada en los siguientes principios: 1) adhesión voluntaria, 2) gestión democrática por parte de los socios, 3) participación económica de los socios, 4) autonomía e independencia, 5) educación, formación e información, 6) intercooperación y 7) interés por la comunidad (Serrano, 2019). Cuando se habla de que esta transición digital de la IA ha de ser necesariamente compatible con la democracia, las formas de organización empresarial basadas precisamente en este principio de la gestión democrática deben jugar un papel protagonista.

Si bien el estudio de la relación entre la economía social y la IA aún no es muy prolífico, existe literatura que analiza la manera en que los principios y valores cooperativos resultan pertinentes para potenciar una IA relacional que supere los efectos adversos en personas y territorios de una IA basada solo en valores capitalistas (Monreal, 2023). La economía social no es ajena a las oportunidades y amenazas de la IA y, para entender estas últimas, se debe hacer referencia a los intereses del sistema capitalista en un contexto de globalización, donde los objetivos de las grandes corporaciones financieras y tecnológicas en la mal llamada economía colaborativa escalan a todos los niveles, también cuando nos referimos a la IA (Sachs, 2020). Ello da como resultado un debilitamiento de las economías y las democracias locales (Sassen, 2014), bajo una ideología económica basada, por un lado, en el crecimiento económico agresivo y, por otro, en un sistema de valores individualistas indiferentes al respeto al medio ambiente o a cualquier otra consideración sobre justicia social (Monreal, 2023).

Para analizar este fenómeno, este artículo se estructura de la siguiente forma: en primer lugar, se ofrecen algunos conceptos clave relacionados con la IA; en segundo lugar, se efectúa un análisis de las potenciales amenazas que pueden surgir, tales como las relacionadas con el desempleo y desigualdades sociales, el poder de las grandes corporaciones, la privacidad y explotación de datos, el deterioro de los sistemas democráticos e inestabilidad social, la discriminación por origen y género o el impacto medioambiental; en tercer lugar, se aborda la regulación de la IA que se está llevando a cabo desde las instituciones públicas, y, por último, se dedica un apartado a las oportunidades desde la economía social en el desarrollo de la IA, para terminar con unas conclusiones.

## Desarrollo de la IA y modelos de negocio

### Conceptos clave de la IA

Definimos la IA como aquel conjunto de sistemas de *software* y *hardware* diseñados por humanos que, ante un objetivo complejo, «actúan en la dimensión física o digital: percibiendo su entorno, a través de la adquisición e interpretación de datos estructurados o no estructurados, razonando sobre el conocimiento, procesando la información derivada de estos datos y decidiendo las mejores acciones para lograr el objetivo dado» (Samoili *et al.*, 2020). En este campo, es también importante definir el concepto de *datificación*, considerado como un «fenómeno contemporáneo que se refiere a la cuantificación de la vida humana a través de información digital, muy a menudo con fines económicos» (Mejías y Couldry, 2022). Otro concepto básico es el de la *plataformización*, que se define como «la penetración de infraestructuras, procesos económicos y marcos gubernamentales de plataformas digitales en diferentes sectores económicos y esferas de la vida, así como la reorganización de las prácticas culturales y los imaginarios en torno a estas plataformas» (Poell *et al.*, 2019).

Por su parte, las «infraestructuras de datos» hacen referencia a aquellos flujos y prácticas materiales, semióticas y sociales que: 1) traducen cosas en números (*datificación*); 2) almacenan, transmiten, analizan y representan datos utilizando lógicas algorítmicas y tecnologías computacionales; 3) integran el uso de datos en una variedad de prácticas; 4) producen nuevos espacios y operaciones de poder a través de prácticas de clasificación, medición y comparación, y 5) generan nuevas prácticas sociales y problematizaciones de lo social, así como nuevas formas de gobernanza (Gulson *et al.*, 2019). Al hablar de *datificación*, es pertinente tener en cuenta el papel que juegan actualmente los activistas del movimiento

de datos abiertos, quienes se encargan de rearticular la noción de democracia, participación y periodismo al aplicar prácticas y valores de la cultura de código abierto a la creación y uso de datos (Baack, 2015). Para este autor (ibídem), las prácticas e ideas de los activistas de datos abiertos son relevantes, ya que conectan *datificación* y cultura de código abierto, ayudando también a entender cómo la *datificación* podría apoyar la acción de públicos y actores no gubernamentales y las grandes empresas. Esta cuestión es especialmente interesante y entronca con dos cuestiones que trataremos más adelante: la regulación de la IA y el papel de las grandes corporaciones.

Van Dijck (2014) aborda el concepto de *dataísmo*, al considerar que los metadatos y los datos se han convertido en una moneda habitual para que los ciudadanos paguen por sus servicios de comunicación y seguridad; una disyuntiva que se ha instalado en la zona de confort de la mayoría de personas. En su trabajo, este autor intenta deconstruir los fundamentos ideológicos de la *datificación*, que tiene sus raíces en afirmaciones ontológicas y epistemológicas problemáticas. Como parte de una lógica más amplia de las redes sociales, la *datificación* muestra características de una creencia secular generalizada, el llamado «dataísmo», que tiene tanto éxito porque multitud de personas –ingenua o inconscientemente– confían su información personal a plataformas corporativas (ibídem).

Otra cuestión cada vez más presente es la referente a la «equidad algorítmica». El hecho de que, cada vez más, la toma de decisiones algorítmica influye en la vida cotidiana de las personas y cómo estos sistemas autónomos pueden causar graves daños a individuos y grupos sociales ha propiciado preocupaciones sobre la equidad (Starke *et al.*, 2022). Es un enfoque centrado en las personas que se viene reclamando cada vez con mayor insistencia y que puede ser una clara oportunidad para la intervención del modelo empresarial de la economía social que, como hemos explicado, sitúa en el centro de su actividad a la persona, por encima del capital.

## **Modelos de negocio en el desarrollo de la IA**

Respecto a casos concretos de grandes corporaciones desarrolladoras de IA, nos detendremos a examinar empresas como LinkedIn y Meta. En el caso de la primera, se utilizan funciones impulsadas por IA que pueden generar perfiles personalizados, con el riesgo de crear perfiles falsos o manipular información personal. También existen sistemas de contratación automatizados que utilizan pruebas de personalidad algorítmicas, las cuales gestionan información obtenida de pruebas psicométricas, que permiten descubrir rasgos de personalidad a través de los cuales es posible predecir un éxito futuro según currículum o perfil

en redes sociales de los usuarios (Rhea *et al.*, 2022). Estamos hablando de herramientas de IA que se utilizan en LinkedIn como CrystalKnows, Dux-Soup, FinalScout, Social Pilot, Mention, Travisnine, Jobscan, People.ai, LinkedIn Talent Insights, Textio, LinkedIn Campaign Manager, Shield, LinkedIn Learning o LinkedIn Analytics (Sharpa Marketing, 2017). Por su parte, Meta, proveedora de Facebook, Instagram o WhatsApp, utiliza una gran variedad de fuentes procedentes de terceros, incluida información disponible públicamente, para así entrenar sus modelos de IA. Y, aunque recientemente ha añadido opciones para que el usuario pueda limitar el uso de su información personal, ello ha sido únicamente para los datos obtenidos a través de terceros, es decir, aquella información que la empresa obtiene a través de otras empresas con las que colabora. Esto quiere decir que el resto de datos generados por los usuarios seguirán disponibles para ser utilizados por parte de Meta para entrenar su IA (Maldita.es, 2023).

Estas empresas se lucran, de esta forma, de los datos que generan los millones de usuarios que tienen alrededor del mundo. Al aceptar la política de privacidad al darse de alta en la plataforma, con un solo clic el usuario muchas veces no es consciente de todo lo que ello comporta. El servicio, a menudo gratuito, siempre precisa de una contraprestación a la que no se otorga toda la importancia que merece: la cesión de los datos emitidos continuamente y de forma dinámica. Meta alcanzó un beneficio neto de 12.369 millones de dólares en el primer trimestre de 2024, más del doble (116,7%) que en el mismo período del ejercicio anterior (Europa Press, 2024).

## Amenazas del desarrollo actual de la IA

Yuval Noah Harari (citado en Barranco, 2024) explica que con la IA el ser humano ha creado «una nueva inteligencia potencialmente más poderosa e inteligente que nosotros. Y si se sale de nuestro control, las consecuencias podrían ser catastróficas». Se abren cientos de escenarios muy peligrosos, porque estamos ante una herramienta que no está en nuestras manos; se están librando millones de nuevos agentes autónomos que podrían volverse más inteligentes que nosotros, por lo que no podemos anticipar ni controlar lo que harán. Además, no estamos hablando únicamente de una tecnología que copia y replica ideas, sino que es capaz de crear ideas propias y también tomar decisiones. Para Harari, «los algoritmos ya han matado a gente»; aunque estos tengan solamente un 1% de responsabilidad, es la primera vez en la historia que las decisiones de la IA sobre qué información difundir han contribuido a un importante cambio histórico (ibídem).

## Desempleo y desigualdades sociales

Si bien la IA ofrece numerosas oportunidades, también aparecen nuevas amenazas, como la robotización del trabajo. Existe literatura que aborda la idea del «reemplazamiento» y la amenaza de ser sustituidos por máquinas. Daniel Innerarity (2022: 31) considera que este sentimiento no sería correcto, ya que «las máquinas reemplazan tareas, pero no ocupan puestos de trabajo, son dos cosas distintas: una cosa es el *task* y otra cosa es el *job*, una cosa es el puesto de trabajo y otra cosa es la tarea». La proliferación de la IA y otras formas de cambio tecnológico que reemplazan a los trabajadores puede ser inequívocamente positiva en una economía óptima en la que los individuos están totalmente asegurados contra cualquier efecto adverso de la innovación, o si ello va acompañado de la forma correcta de

**Las cooperativas diluyen el conflicto entre el capital y el trabajo, ya que la condición de socio cooperativista sitúa en un mismo sujeto la percepción de los rendimientos derivados del trabajo y los rendimientos del capital.**

redistribución. Pero el cambio tecnológico que reemplaza a los trabajadores puede no solo conducir a estos a obtener una fracción cada vez menor del ingreso nacional, sino que puede empeorar su situación en términos absolutos (Korinek y Stiglitz, 2019).

La robotización tendría como consecuencia un progresivo «crecimiento masivo del desempleo, creciente desigualdad y mayor crecimiento de la productividad a través de la automatización» (Gries y Naudé, 2018). Y, aunque aún no se dispone de evidencia científica suficiente para aseverar con seguridad que la IA causará pérdidas masivas de empleos, es cierto que, a la vez que se está produciendo un crecimiento más rápido de la productividad y el PIB, el crecimiento de los salarios y la productividad se están estancando, mientras la desigualdad aumenta. Aquí es donde aparece una de las claves: «si los ingresos laborales no se benefician de las ganancias económicas generadas por el progreso en IA, el consumo puede estancarse y restringir el crecimiento» (ibídem).

En este contexto, los sistemas algorítmicos y la IA juegan un papel fundamental en la distribución de bienes y resultados; por eso, pueden tener una afectación importante en la reproducción de desigualdad en el futuro (Zajko, 2022). La equidad, como elemento central del desarrollo sostenible, no puede desligarse de las formas y modelos que generan la dispersión en la posesión o percepción de la riqueza (Serrano, 2024). La literatura científica no es muy prolífica en analizar las raíces microeconómicas o el papel del modelo empresarial en la generación de esas desigualdades, por lo que, en este sentido, tampoco se identifica la economía social, las cooperativas y las sociedades laborales como agentes correctores que puedan contribuir a reducir la desigualdad. Podemos afirmar que las cooperativas diluyen el conflicto entre el capital y el trabajo, ya que la condición de socio

cooperativista sitúa en un mismo sujeto la percepción de los rendimientos derivados del trabajo y los rendimientos del capital (doble condición trabajador-propietario) (ibídem). Piketty (2014) hace referencia a la cultura emprendedora de nuestro tiempo, y aquí hay que pensar inevitablemente en Silicon Valley y en diferentes empresarios concretos cuyas fortunas siguen aumentando año tras año, junto con la «tendencia  $r > g$ », esto es, la tasa de retorno sobre el patrimonio es mayor que la tasa de crecimiento, lo que nos lleva a un incremento automático del capital inicial, que provocará que las fortunas crezcan y se perpetúen hasta límites que sobrepasen la utilidad social (Lobato, 2014).

Además de la eventual eliminación de tareas que hoy desarrollan personas a causa del desarrollo de la IA, otro de los grandes riesgos es la *plataformización* de las empresas, ya que los trabajadores actuarían como falsos autónomos que facturan sus servicios a la plataforma en cuestión, si bien con una relación laboral exclusiva y permanente. La Confederación Europea de Sindicatos (CES) calcula que actualmente hay 28 millones de personas en Europa en esta situación: personal laboral sin capacidad de negociar de forma colectiva sus condiciones de trabajo (Cruz, 2023).

## El poder de las grandes corporaciones

Nos encontramos ante un futuro marcado por la presencia de una lógica económica sin precedentes que aspira a dominar la siguiente modernidad, con origen en Silicon Valley (Zuboff, 2020). Convivimos con la cara oculta de las prácticas de estas grandes empresas sin reparar en ellas por su cualidad etérea, imperceptibles a simple vista. Existe un «capitalismo de la vigilancia» (ibídem), enfocado hacia un futuro gobernado desde la alegalidad por Google, Facebook, Amazon y otras grandes empresas de internet e IA que explotan la experiencia humana sin consentimiento y vulneran el derecho a la privacidad del individuo (Moscardó, 2021). El papel preponderante de las grandes corporaciones de Silicon Valley puede comportar determinadas amenazas sociales: intensificación de la desigualdad y la jerarquización social, mayor exclusión, usurpación de derechos y eliminación de la vida personal en todo aquello que la hace justamente personal. Los actos de desposesión digital que cometen los capitalistas de la vigilancia «imponen una nueva forma de control sobre individuos, poblaciones y sociedades enteras» (Zuboff, 2020). Todo ello tiene su base en la posición envidiable de las grandes empresas tecnológicas, que durante 20 años han utilizado todo tipo de fórmulas de extracción de datos a bajo precio, de manera que hoy en día pocas instituciones, incluidas las administraciones públicas, pueden competir con ellas, a la vez que han recibido cuantiosas ayudas públicas (Morozov, 2018).

El «capitalismo ahora está muerto» y su dinámica ya no rige nuestras economías, ya que su papel ha sido reemplazado por algo fundamentalmente diferente, el «tecnofeudalismo» (Varoufakis, 2023). Lo que ha matado al capitalismo es el capital mismo, una mutación causada por: 1) la privatización de Internet por parte de las grandes tecnológicas de Estados Unidos y China, y 2) la manera en que Occidente, los gobiernos y los bancos centrales respondieron a la gran crisis de 2008. Todo ello habría demolido los dos grandes pilares del capitalismo: los mercados y las ganancias. Así, los mercados han sido reemplazados por las plataformas digitales comerciales, que son feudos; por lo que respecta a las ganancias, estas han sido sustituidas por su predecesor feudal, el alquiler (ibídem). El término «tecnofeudalismo» ya había sido utilizado anteriormente por otros autores. Para Durand (2021a), las plataformas están en vías de convertirse en feudos, y quien controla estas infraestructuras concentra un poder, tanto político como económico, sobre quienes están vinculados a ellas. Nos encontraríamos, así, en una fase senil del capitalismo donde existiría una captura de rentas por parte de estas plataformas y grandes empresas digitales (Cairó, 2022; Durand, 2021b).

Si nos detenemos en la ética empresarial, también existen amenazas tales como, entre otras, la falta de transparencia y aplicabilidad de los sistemas de IA, la ausencia de rendición de cuentas, las prácticas manipuladoras de marketing o los mecanismos que inducen a la adicción (Masut y Schmidt, 2024).

## **Privacidad y explotación de datos**

Existen las amenazas de las tecnologías de IA en lo correspondiente al derecho a la privacidad como, por ejemplo, la existencia de un ecosistema en expansión de recopilación e intercambio de datos personales, en gran medida no transparentes. Estos sistemas afectan a los enfoques gubernamentales en materia de vigilancia y administración de justicia, determinan la accesibilidad de los servicios públicos, deciden quién tiene posibilidades de ser contratado para un puesto de trabajo y afectan a la información que las personas ven y pueden compartir en línea, por ejemplo. Además, se produce un claro riesgo de discriminación vinculado a las decisiones basadas en la IA (Naciones Unidas, 2021), pudiendo existir acciones perjudiciales tales como perfilamientos, vigilancia, ataques cibernéticos y afectaciones a la capacidad de decisión de los propios usuarios. Por ejemplo, cuando estos comparten sus datos con determinadas empresas a través de Internet, muchas veces no saben si la información que están facilitando será vendida a terceros o si los sistemas de IA que recolectan dicha información están cifrados o encriptados (Murrugarra, 2024).

La IA permite la recopilación de grandes cantidades de datos personales de todo tipo (búsquedas en la red, hábitos de consumo, etc.), que permite la creación de perfiles de usuarios detallados, lo que representa una amenaza para la privacidad. Al respecto, es tan importante la regulación por parte de los poderes públicos como la implementación de prácticas éticas por parte de las empresas, así como una concienciación mayor de los usuarios a la hora de ceder sus datos a estas empresas (Alfonso *et al.*, 2023).

## Deterioro de los sistemas democráticos e inestabilidad social

Aguirre (2021) advierte de que «las tecnologías digitales, y en concreto la IA, lejos de mejorar la democracia y la participación ciudadana, algunos de sus usos han conducido a la denigración de la deliberación democrática». Uno de los desafíos éticos principales son la falsedad y la saturación de información, que fomentan la promoción de emociones concretas e interesadas, carentes de transparencia, lo que puede condicionar la deliberación democrática necesaria para votar, promover iniciativas legislativas, etc. Esto lleva a la *posverdad* e *infoxicación*, que pueden comportar que la ciudadanía confunda realidad virtual con vida real colectiva (ibídem). Por ello, la teoría deliberativa de la democracia ofrece «el mejor marco en el que organizar la conversación para corregir la supuesta objetividad algorítmica con las plurales preferencias de los humanos» (Innerarity, 2023a). Según Harari, en Estados Unidos se dispone de la tecnología de la información más sofisticada de la historia, pero sus ciudadanos están perdiendo la capacidad de hablar entre sí y escucharse. Y aquí la sospecha es que esta tecnología, en lugar de facilitar la comunicación, la dificulta, porque además existen muchas voces no humanas; es decir, muchas entidades tecnológicas han entrado en el debate; por ejemplo, se calcula que más del 20% del contenido de X (antiguo Twitter) lo difunden *bots*. Esto es, los algoritmos que deciden qué voces amplificar y silenciar no son humanos (Harari, citado en Barranco, 2024). En este contexto, el papel de los medios de comunicación es fundamental a la hora de definir el proceso deliberativo en los sistemas democráticos. Nikunen y Hokka (2019) estudian cómo los medios de servicio público —que consideran parte del Estado de bienestar— y sus valores de universalidad, igualdad, diversidad y calidad se ven afectados por la *datificación* y un entorno de medios de comunicación de plataforma. Inquieta que sean los algoritmos los enterradores de la democracia.

Innerarity (2022) se pregunta si estamos entrando en un nuevo totalitarismo de la mano de la ideología de la optimización, y si siguen teniendo sentido la información razonada, la decisión propia y el autogobierno democrático en esos nuevos entornos tecnológicos. Así, hay quien denuncia un *big nudging*: gigantesco sistema de información superinteligente, central y tecnocrático, que

representaría una forma de dictadura en la medida en que vaciaría de contenido decisonal a nuestras organizaciones políticas (Helbing *et al.*, 2017). Por ello, «no deberíamos minusvalorar el riesgo de que el tecnoautoritarismo resulte cada vez más atractivo en un mundo en el que la política cosecha un largo listado de fracasos» (ibídem, 2020). Este es uno de los grandes peligros: amplias capas de la sociedad achacan los fracasos de nuestras sociedades al funcionamiento de las democracias liberales, y eso puede comportar la seducción de un discurso basado en una supuesta solución mágica a todos nuestros problemas por parte de las herramientas que nos ofrece la IA, también en lo concerniente a la política. Innerarity (2023b) explica que «en poco tiempo hemos pasado del ciberentusiasmo a la “tecnopreocupación”, aunque los datos conducen a un conocimiento que seguirá siendo rebatible y humano. Demasiado humano». En política, las decisiones deben tomarlas las personas y no las máquinas (Innerarity, 2022). La falta de información, la tergiversación de noticias y las fuentes carentes de fiabilidad pueden suponer un sesgo claramente perjudicial a la hora de tomar decisiones.

## **Discriminación por origen y género**

La amenaza de la discriminación por cuestión de origen o género es también capital, ya que la IA se nutre de datos generados en una sociedad históricamente discriminatoria. Para luchar contra esta situación, se han propuesto varias técnicas para lograr la imparcialidad algorítmica: el llamado «algoritmo sin restricciones» maximiza la seguridad pública y al mismo tiempo satisface un concepto importante de igualdad, es decir, que todos los individuos estén sujetos al mismo estándar, independientemente de su origen (Corbett-Davies *et al.* 2017). Respecto al género, los datos que los alimentan y los resultados de la toma de decisiones algorítmicas pueden estar sesgados, lo que potencialmente perjudica a las mujeres y perpetúa la discriminación (Gutiérrez, 2021). Así, el sesgo algorítmico de género se puede definir como el cálculo algorítmico basado en datos con una discriminación o invisibilización de las mujeres, creando resultados adversos para ellas. Además, la discriminación de género en la vida real incorporada en los datos puede dar lugar a sesgos en el mundo digital (ibídem). Por todo ello, los datos históricos no pueden ser la única base para la política algorítmica que afecta a las mujeres y, además, como propuesta, se deberían incluir más mujeres como diseñadoras de algoritmos y también como evaluadoras (ibídem). Como demuestran Zhao *et al.* (2017), los modelos de predicción estructurados amplifican el sesgo y hay que proponer métodos para reducir este efecto, pero quedan importantes retos para el trabajo futuro. Porque los algoritmos de aprendizaje automático se utilizan en contextos sensibles que afectan el curso de las vidas humanas, como los préstamos de crédito o la justicia penal, por ejemplo, y

se sabe que las máquinas «objetivas» basan sus decisiones únicamente en hechos y no se ven afectadas por los sesgos cognitivos humanos, las tendencias discriminatorias o las emociones (Tolan, 2019), y se trata de una cuestión que debe ser tenida en cuenta para no perpetuar discriminaciones de forma injusta.

## Impacto medioambiental

También hay que hacer referencia a las amenazas relacionadas con los recursos naturales: consumo energético de los grandes centros de datos, extracción de agua para la tecnología de IA (refrigeración de servidores, etc.), materias primas, generación de residuos (*hardware* que queda obsoleto rápidamente, liberación al medio ambiente de sustancias como plomo, mercurio y cadmio, etc.) (Masut y Schmidt, 2024). De la misma forma, existe la amenaza de impactar negativamente en el respeto a los derechos humanos: a la vida privada y la intimidad, a la no discriminación, al derecho a la vida, a la libertad y a la seguridad personal, a la libertad de opinión y de expresión, a la protección de la propiedad intelectual, derechos laborales, etc. (ibídem). Uno de los pilares para lograr la transición energética es la transición digital, pero la tecnología digital también requiere de altos consumos de energía (Terrones, citado en Coll, 2024).

Tabla 1. Principales amenazas del desarrollo actual de la IA

Eliminación de puestos de trabajo
Aumento de la precariedad laboral
Crecimiento de las desigualdades sociales
Preponderancia de las grandes corporaciones
Plataformización de las empresas
Distorsión de la competencia entre empresas
Impacto en el medio ambiente
Difusión de la desinformación
Denigración de la deliberación democrática
Perjuicio en los derechos fundamentales
Aumento de la discriminación
Riesgos de seguridad informática
Afectación a la privacidad
Falta de transparencia y rendición de cuentas
Deterioro de los sistemas democráticos
Inestabilidad social

Fuente: elaboración propia.

## Regulación de la IA

Existe la necesidad de regular y controlar la IA, teniendo en cuenta los procesos de rendición de cuentas (Prabhakaran *et al.*, 2022), ya que los riesgos que entraña su progreso son evidentes y la inquietud social puede ir en aumento. Un ejemplo de ello es el recientemente aprobado Reglamento de IA de la UE del 13 de junio de 2024. Tal y como se contempla en su introducción, su objetivo es «mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico uniforme, en particular para el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial (...) en la Unión, de conformidad con los valores de la Unión, a fin de promover la adopción de una IA centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la Unión Europea (...), incluidos la democracia, el Estado de derecho y la protección del medio ambiente, proteger frente a los efectos perjudiciales de los sistemas de IA en la Unión, así como brindar apoyo a la innovación» (Unión Europea, 2024). Hay que aclarar que esta normativa se aplica a los proveedores de sistemas de IA en el mercado de la Unión Europea (UE), o quienes los ponen en servicio en este territorio; por lo que se puede hablar de impacto global. Sin embargo, no se aplica a los sistemas de IA desarrollados en la Unión y exportados al exterior. Además, sectores como la aviación civil y la seguridad nacional quedan fuera de este reglamento (Masut y Schmidt, 2024).

Es necesario recordar que, ya previamente, la Comisión Europea había emitido en 2018 una serie de recomendaciones basadas en siete ejes para el desarrollo de las aplicaciones en IA: 1) que la IA no sea autónoma, sino supervisada por especialistas o profesionales; 2) que sea robusta y segura; 3) que respete la privacidad y no haga un uso ilícito de los datos personales o sensibles; 4) que sea transparente; 5) que garantice resultados equitativos y sin sesgos que vulneren los derechos de grupos de personas concretos (mujeres, pobres, ancianos, etc.) ; 6) que esté dirigida al bienestar y que sea sostenible; y 7) que rinda cuentas (Gibert, 2024). En este sentido, el citado reglamento de la UE representa un avance importante en la regulación internacional de la IA. El problema es que son las grandes empresas tecnológicas las que están definiendo el relato, introduciendo sus aplicaciones en ámbitos centrales de nuestra vida; además, la regulación sobre los sistemas de IA se está haciendo a posteriori de un sistema demasiado permisivo (Vallès, entrevistada en Toro, 2024). La ley siempre irá por detrás de la tecnología y es prácticamente seguro que las regulaciones presentes y futuras casi siempre llegarán tarde (Gibert, 2024).

## Oportunidades desde la economía social en el desarrollo de la IA

La IA, sin embargo, también ofrece oportunidades, y es probable que la necesitemos para sobrevivir como especie y abordar los retos del siglo XXI: el cambio climático, el transporte eficiente y autónomo, la medicina de precisión o personalizada, entre otros. Todos los grandes descubrimientos científicos solo se podrán llevar a cabo analizando una gran cantidad de datos y, para ello, es necesaria la IA (Oliver, entrevistada en Salicrú, 2024). De la misma forma, para las empresas, la IA puede permitir desarrollar una nueva generación de productos y servicios, particularmente en sectores como la economía verde y circular, maquinaria, agricultura, atención sanitaria, moda, turismo, etc.; puede impulsar las ventas, mejorar el mantenimiento de las máquinas, aumentar la producción y la calidad, mejorar el servicio al cliente y ahorrar energía, etc. Igualmente, la IA que se utiliza en los servicios públicos puede reducir los costes y ofrecer nuevas posibilidades en el transporte público, la educación, la energía y la gestión de residuos, por ejemplo (Parlamento Europeo, 2020).

**La economía social puede representar una oportunidad para la democratización de la IA. Asimismo, un funcionamiento de la IA basado en sus valores, minimizaría los riesgos y amenazas que existen a la hora de repartir el beneficio que representa la IA a nivel empresarial.**

En este contexto, las empresas de economía social no pueden a renunciar a la IA, pues esta es muy útil en diferentes áreas y tareas: innovación, producción, logística, gestión de personas, etc. La clave está en hacer un uso de estas herramientas basado en los siete principios fundamentales de la economía social enumerados al inicio del presente artículo. Así, las organizaciones que desarrollen aplicaciones y herramientas de IA deberán priorizar los intereses de las personas, basándose en unos valores éticos y sociales y prestando especial atención a la discriminación de las personas y los colectivos, sobre todo en lo concerniente a los puestos de trabajo. En resumen, IA y economía social pueden y deben relacionarse para promover un desarrollo de la IA de potente oportunidad tecnológica, basado en principios como la justicia, la equidad o la sostenibilidad. De hecho, se considera que existe la necesidad de desarrollar la IA de una forma alternativa. Stiglitz (citado en Co-operative News, 2016) ha defendido que las cooperativas, junto con otras iniciativas de la economía social, proporcionan «un tercer pilar clave para remodelar la economía globalizada». Por su parte, Rifkin (citado en Alianza Cooperativa Internacional, 2016) explica que las cooperativas podrían liderar un cambio de paradigma en diferentes sectores, también en el tecnológico,

y las alienta a utilizar su dominio web «.coop» y a crear nuevos algoritmos y aplicaciones para utilizar los datos de forma ética, a la vez que incita a los cooperativistas a desarrollar sus propias redes sociales, tipo Facebook, Google, Twitter, o plataformas de transporte.

La economía social puede representar una oportunidad para la democratización de la IA. Con la Cuarta Revolución Industrial, al igual que en las anteriores con el surgimiento de cooperativas, mutuas y asociaciones, se configura un nuevo sector tecnológico con rasgos distintivos al que denominamos «economía digital sostenible» (Pastor, 2022), referente a la idea de que, ante la progresiva crisis de transición ecológica, se necesita algo más que una conciencia social para enfrentarla. Aquí, las cooperativas de datos representan una oportunidad organizativa para empoderar a los individuos y darles el control sobre los datos que comparten, monetizándolos o entregándolos para la investigación. Como ejemplo, encontraríamos Driver's Seat, una cooperativa de conductores estadounidense que recopila datos relacionados con su trabajo a partir de los teléfonos inteligentes de los conductores de viajes compartidos; o la irlandesa Resonate, que ofrece el primer servicio de transmisión de música propiedad de la comunidad de músicos; o la española Salus.coop, una cooperativa de donantes de datos para la investigación para el Bien Común (ibídem).

Si en el apartado de las amenazas del desarrollo de la IA hacíamos referencia a los medios de comunicación, aquí estos aparecen como una nueva oportunidad, ya que existe la opción de implementar modelos de economía social en medios de comunicación, constituyendo cooperativas de periodistas, por ejemplo. La crisis del sector ha llevado a reducciones de plantillas en medios de comunicación tradicionales, y existen ejemplos de periodistas que, al perder su puesto de trabajo, han optado por constituir empresas de economía social a fin de seguir desarrollando su actividad profesional, creando nuevos medios basados en principios de gestión democrática, ética, etc. Zia (2023) afirma que la democratización de la IA consiste en ponerla al alcance de todos, independientemente de su procedencia o recursos. El objetivo clave es ofrecer a todo el mundo las mismas oportunidades de beneficiarse de la IA para «potenciar la innovación y fomentar la creatividad». Para este autor, ello implica aspectos como la accesibilidad, la educación y formación, la colaboración y apertura o las consideraciones éticas. Así, existirían diferentes tipos de democratización de la IA: de datos, de algoritmos, de informática y del conocimiento. Las ventajas de la democratización de la IA son múltiples: reducción de la escasez de competencias, eficiencia de costes, consecución de una mayor productividad, mejora de la innovación empresarial, así como la eliminación de barreras, es decir, personas de todo el mundo, independientemente de su situación financiera, podrían iniciar su andadura en la IA (ibídem). Aquí, nuevamente, el papel de la economía social puede ser importante.

En la democratización de la IA sería pertinente plantear la consecución de un desarrollo de sus herramientas centrado en las personas y sus intereses. Ello no debe ir reñido necesariamente con la pérdida de eficiencia o incluso de generación de beneficios económicos, pues hay numerosos ejemplos de empresas de economía social a nivel global que, a la vez que sitúan en el centro de su actividad a las personas, son capaces de funcionar de forma eficiente y generar beneficios económicos que repercutirán positivamente en sus socios y también en las comunidades y sociedades en las que actúan. Un funcionamiento de la IA basado en los valores de la economía social, sin duda alguna minimizaría los riesgos y amenazas que existen a la hora de repartir el beneficio que representa la IA a nivel empresarial. El papel de las administraciones públicas, en este sentido, se antoja fundamental, tanto en lo que respecta a la regulación de la IA como en lo correspondiente al fomento y soporte de iniciativas empresariales que respeten los derechos de las personas y el medio ambiente.

Tabla 2. Oportunidades de la economía social en el desarrollo de la IA

Gestión democrática de la empresa
Preservación de los puestos de trabajo
Empleo de calidad
Resultados orientados al bien común
Reparto equitativo de beneficios
Igualdad de oportunidades
Establecimiento de sinergias con otras empresas – Intercooperación
Fomento de la creación de pequeñas y medianas empresas
Información transparente, verificable y fiable
Valores éticos y sociales en el desarrollo de la actividad
Respeto por los derechos humanos
Rendición de cuentas
Respeto al medio ambiente y las comunidades locales
Respeto de la legalidad
Respecto de la privacidad del usuario
Mejora de la equidad algorítmica
Mejora de la transparencia
Compromiso democrático

Fuente: elaboración propia.

## Conclusiones

La democratización de la IA significa, en primer lugar, que esté al alcance de todo el mundo, independientemente de sus antecedentes o recursos (Zia, 2023). Asimismo, es fundamental que el desarrollo de la IA no implique una acumulación de riqueza en pocas manos, en ese «capitalismo de la vigilancia» (Zuboff, 2020) que avanza hacia un futuro disruptivo gobernado desde la alegalidad por grandes corporaciones como Google, Facebook o Amazon, las cuales, a su vez, se benefician de recursos públicos desde instituciones muchas de las cuales son democráticas. Si el desarrollo de la IA, en constante evolución y progreso a un ritmo

**Es fundamental que el desarrollo de la IA no implique una acumulación de riqueza en pocas manos, en ese «capitalismo de la vigilancia» del que habla Zuboff, que avanza hacia un futuro disruptivo gobernado desde la alegalidad por grandes corporaciones.**

vertiginoso, se deja en manos de tan solo unas pocas empresas, el beneficio económico generado repercutirá en una minoría, aumentando las desigualdades sociales y excluyendo a las capas sociales más desfavorecidas. Asimismo, dado que la robotización implicará inevitablemente la eliminación de tareas que tradicionalmente

desarrollaban personas, se debería garantizar, desde las instituciones públicas, que el modelo «tecnofeudalista» (Varoufakis, 2023) en el que nos encontramos no continúe beneficiando exclusivamente a unos pocos privilegiados.

La regulación de la IA resulta imprescindible ante las eventuales amenazas actuales. Lo cierto es que existe un desequilibrio en el desarrollo de la IA y su marco regulatorio a nivel internacional. Todo ello supone que los ordenamientos jurídicos nacionales queden comprometidos y haya que replantear los mecanismos de gobernanza, adoptando sistemas multinivel. Este artículo ha hecho referencia al Reglamento de IA de la UE recientemente aprobado, que representa un gran paso adelante para evitar esta amenaza, además de constituir un punto de partida que podría inspirar a otros países u organizaciones internacionales, al basarse en los valores esenciales de la UE, tales como la protección de la salud, la seguridad y los derechos fundamentales, así como la democracia, el Estado de derecho, la protección del medio ambiente, entre otros. La coordinación, negociación y acuerdo entre países y organizaciones internacionales será fundamental si se pretende democratizar la IA.

En este contexto, el modelo empresarial de la economía social representa una oportunidad para luchar contra las amenazas que comporta el desarrollo de la IA. Al situar la persona, y no el capital, en el centro de la actividad empresarial, la defensa y preservación del puesto de trabajo de cada socio trabajador adquiere una relevancia primordial, y se consigue así un reparto más equitativo de la

riqueza que se genere. En consecuencia, los beneficios que se pudieran obtener gracias a la IA quedarían más equitativamente repartidos y, por ende, se colaboraría en la reducción de la desigualdad a escala global. Si tenemos en cuenta que la economía social se encuentra presente en los diferentes sectores económicos, no existe razón alguna para pensar que no pueda estarlo también en el desarrollo de la IA, a través de empresas que sitúen en un primer plano valores fundamentales como la gestión y participación democrática, la autonomía e independencia, así como el interés por la comunidad.

## Anexo

Referencia bibliográfica	Ideas clave
<b>Aguirre, 2021</b>	Algunos de los usos de las tecnologías digitales, y en concreto la IA, han conducido a la denigración de la deliberación democrática.
<b>Cruz, 2023</b>	La individualización de las relaciones laborales es uno de los riesgos más grandes de la IA.
<b>Durand, 2021a y 2021b</b>	Nos encontraríamos en una fase senil del capitalismo con una captura de rentas por parte de las plataformas y grandes empresas digitales.
<b>Gibert, 2024</b>	La ley siempre irá por detrás de la tecnología y es prácticamente seguro que las regulaciones presentes y futuras casi siempre irán tarde.
<b>Gries y Naudé, 2018</b>	Se hace referencia a cuestiones como un eventual crecimiento masivo del desempleo, creciente desigualdad y mayor crecimiento de la productividad a través de la automatización.
<b>Guiérrez, 2021</b>	Los datos y los resultados de la toma de decisiones algorítmicas pueden estar sesgados, lo que potencialmente perjudica a las mujeres y perpetúa la discriminación.
<b>Harari, 2024</b>	Con la IA el ser humano ha creado una nueva inteligencia potencialmente más poderosa e inteligente que nosotros, y si se sale de nuestro control, las consecuencias podrían ser catastróficas.
<b>Innerarity, 2020</b>	En poco tiempo hemos pasado del <i>ciberentusiasmo</i> a la <i>tecnopreocupación</i> ; en vez de entender las nuevas tecnologías como fuentes de capacitación, cada vez las consideramos más como artefactos para el <i>desempoderamiento</i> . No deberíamos minusvalorar el riesgo de que el <i>tecnautoritarismo</i> resulte cada vez más atractivo en un mundo en el que la política cosecha un largo listado de fracasos.
<b>Innerarity, 2022</b>	Las máquinas reemplazan tareas pero no ocupan puestos de trabajo, son dos cosas distintas: una cosa es el <i>task</i> y otra cosa es el <i>job</i> .
<b>Innerarity, 2022</b>	En política las decisiones deben tomarlas las personas, y no las máquinas.
<b>Innerarity, 2023b</b>	Mientras los sistemas humanos sean complejos, contradictorios y paradójicos, los datos conducirán a conocimiento que seguirá siendo rebatible y humano. Demasiado humano.
<b>Korinek y Stiglitz, 2019</b>	La IA y otros cambios en la tecnología requieren grandes ajustes. Cuanto más dispuesta esté la sociedad a apoyar la transición necesaria y a prestar apoyo a aquellos que se quedan atrás, más rápido será el ritmo de innovación que la sociedad pueda acomodar.

<b>Korinek y Stiglitz, 2019</b>	La IA y otros cambios en la tecnología requieren grandes ajustes. Cuanto más dispuesta esté la sociedad a apoyar la transición necesaria y a prestar apoyo a aquellos que se quedan atrás, más rápido será el ritmo de innovación que la sociedad pueda acomodar.
<b>Monreal, 2023</b>	Los principios y valores cooperativos de la economía social resultan más pertinentes para potenciar una inteligencia racional que supere o mitigue los efectos adversos en personas y territorios de una IA basada en valores capitalistas.
<b>Morozov, 2018</b>	Las grandes posibilidades de Internet pueden utilizarse para hacer campaña a favor de las libertades, pero también para todo lo contrario: fomentar la represión, el conformismo y la ceguera.
<b>Pastor, 2022</b>	La economía social digital podría cambiar la forma en que se valoran los datos, los activos y los recursos laborales y comercializados, empoderando e incentivando a individuos, empresas y gobiernos a cooperar liberando valor financiero de cosas que actualmente se desperdician o se descartan, tratados como económicamente invaluables.
<b>Piketty, 2014</b>	En el sector de las grandes empresas tecnológicas, determinadas fortunas siguen aumentando año tras año, y junto con la tendencia $r > g$ , nos lleva a un incremento automático del capital inicial que provocará que las fortunas crezcan y se perpetúen hasta límites que sobrepasen la utilidad social.
<b>Rifkin, 2016</b>	Las cooperativas podrían liderar un cambio de paradigma en diferentes sectores.
<b>Starke et al., 2022</b>	La toma de decisiones algorítmica influye cada vez más en la vida cotidiana de las personas.
<b>Stiglitz, 2016</b>	Las cooperativas, junto con otras iniciativas de la economía social, proporcionan un tercer pilar clave para remodelar la economía globalizada.
<b>Van Dijk, 2014</b>	Los metadatos y los datos se han convertido en una moneda habitual para que los ciudadanos paguen por sus servicios de comunicación y seguridad.
<b>Varoufakis, 2023</b>	El capitalismo ahora está muerto y su dinámica ya no rige nuestras economías, ya que su papel ha sido reemplazado por algo fundamentalmente diferente: el <i>tecnofeudalismo</i> .
<b>Zajko, 2022</b>	Los sistemas algorítmicos y la IA juegan un papel fundamental en la distribución de bienes y resultados, y pueden tener una afectación importante en la reproducción de desigualdad en el futuro.
<b>Zia, 2023</b>	La democratización de la IA consiste en ponerla al alcance de todos, independientemente de su procedencia o recursos. El objetivo clave es ofrecer a todos las mismas oportunidades de beneficiarse de la IA para potenciar la innovación y fomentar la creatividad.
<b>Zuboff, 2020</b>	Estamos ante un futuro marcado por la presencia, en modo oculto, de una lógica económica sin precedentes que aspira a dominar la siguiente modernidad, con origen en Silicon Valley, especialmente la empresa Google.

Fuente: elaboración propia.

## Referencias bibliográficas

- Aguirre Sala, Jorge Francisco. «Los desafíos de la transformación digital de la democracia». *RECERCA. Revista de Pensament i Anàlisi*, vol. 26, n.º 2 (2021). DOI: <https://doi.org/10.6035/recerca.4660> (en línea).
- Alfonso Rojas, Juan Ricardo; Díaz, Sthefany Loriet y Ramírez Montúfar, Álvaro Hernando. «Efecto de la monetización de datos personales, provenientes de plataformas digitales, sobre el derecho a la privacidad». *Lex, Revista de Investigación en Ciencias Jurídicas*, vol. 6, n.º 22 (2023) (en línea) <http://repositorio.cidcuador.org/jspui/handle/123456789/2859>
- Alianza Cooperativa Internacional. «Jeremy Rifkin: Los modelos cooperativos podrían liderar la revolución digital», (27 de octubre de 2016) (en línea) <https://ica.coop/es/sala-de-prensa/noticias/jeremy-rifkin-modelos-cooperativos-podrian-liderar-revolucion-digital>
- Baack, Stefan (2015). «Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism». *Big Data & Society*, vol. 2, n.º 2 (2015), 2. DOI: <https://doi.org/10.1177/2053951715594634>
- Barranco, Justo. «Yuval Noah Harari “Los algoritmos ya han matado a gente”». *La Vanguardia*, (3 de septiembre de 2024) (en línea) <https://www.lavanguardia.com/cultura/20240903/9907895/yuval-noah-harari-nexus-sapiens-inteligencia-artificial-trump-israel-palestina.html>
- Cairó Céspedes, Gemma. «Reseña: Cédric Durand; Tecnofeudalismo. Crítica a la economía digital (2021)». *Revista de Economía Crítica*, n.º 33 (2022), p. 102-104.
- Chávez Ávila, Rafael y Monzon Campos, José Luis. «La economía social ante los paradigmas económicos emergentes: innovación social, economía colaborativa, economía circular, responsabilidad social empresarial, economía del bien común, empresa social y economía solidaria». *CIRIEC-España, Revista de Economía Pública, Social y Cooperativa*, n.º 93 (2018), p. 5-50. DOI: 10.7203/CIRIEC-E.93.12901
- Coll, Maria. «Antonio Luis Terrones: “Estem davant d’un blanquejament ètic de la IA”». *Revista Valors*, (2 de marzo de 2024) (en línea) <https://valors.org/estem-davant-dun-blanquejament-etic-de-la-ia/>
- Co-operative News. «Joseph Stiglitz proposes co-op models as an alternative to trickle-down economics». *Coop News*, (12 de octubre de 2016) (en línea) <https://www.thenews.coop/110090/sector/community/joseph-stiglitz-proposes-co-ops-models-alternative-trickle-economics/>
- Corbett-Davies, Sam; Pierson, Emma; Feller, Avi; Goel, Sharad y Huq, Aziz. «Algorithmic decision making and the cost of fairness». *ArXiv*, (2017), p. 797-806 (en línea) <https://arxiv.org/abs/1701.08230>

- Cruz Fuentes, Daniel. «El sindicalismo es esencial para humanizar la inteligencia artificial». *Revista Perspectiva*, (11 de abril de 2023) (en línea) [https://perspectiva.fsc.ccoo.es/noticia:660887--El\\_sindicalismo\\_es\\_esencial\\_para\\_humanizar\\_la\\_inteligencia\\_artificial](https://perspectiva.fsc.ccoo.es/noticia:660887--El_sindicalismo_es_esencial_para_humanizar_la_inteligencia_artificial)
- Durand, Cédric. «Tecnofeudalismo: la nueva gleba digital». *Viento Sur*, n.º 173, (29 de enero de 2021a) (en línea) <https://vientosur.info/tecnofeudalismo-la-nueva-gleba-digital/>
- Durand, Cédric. *Tecnofeudalismo. Crítica a la economía digital*. Donostia: Kaxilda, 2021b.
- Europa Press. «Meta más que duplica sus beneficios en el primer trimestre y eleva previsiones de gasto por la IA», (25 de abril de 2024) (en línea) <https://www.europapress.es/economia/noticia-meta-mas-duplica-beneficios-primer-trimestre-eleva-previsiones-gasto-ia-20240425133550.html>
- Gibert, Karina. «L'ètica de la IA: més enllà de la regulació». *Revista Valors*, (13 de marzo de 2024) (en línea) <https://valors.org/letica-de-la-ia-mes-enlla-de-la-regulacio/>
- Gries, Thomas y Naudé, Wim. «Artificial Intelligence, Jobs, Inequality and Productivity: Does Aggregate Demand Matter». *IZA Discussion Paper Series*, n.º 12005 (noviembre de 2018) (en línea) <https://www.iza.org/publications/dp/12005/artificial-intelligence-jobs-inequality-and-productivity-does-aggregate-demand-matter>
- Gulson, Kalervo N. y Sellar, Sam. «Emerging data infrastructures and the new topologies of education policy». *Environment and Planning D: Society and Space*, vol. 37, no.º 2 (2019), p. 350-366. DOI: <https://doi.org/10.1177/0263775818813144>
- Gutiérrez Almazor, Miren. «Algorithmic Gender Bias and Audiovisual Data: A Research Agenda». *International Journal of Communication*, n.º 15 (2021), p. 439-461.
- Harari, Yuval Noah. *Nexus: Una breve historia de las redes de información desde la Edad de Piedra hasta la IA*. Barcelona: Debate, 2024.
- Helbing, Dirk; Frey, Bruno S.; Gigerenzer, Gerd; Hafen, Ernst; Hagnet, Michael; Hofstetter, Yvonne; Van den Hoven, Jeroen; Zicari, Roberto V. y Zwitter, Andrej. «Will Democracy Survive Big Data and Artificial Intelligence?». *Scientific American*, (25 de febrero de 2017) (en línea) <https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/>
- Innerarity, Daniel. «El impacto de la inteligencia artificial en la democracia». *Revista de las Cortes Generales*, n.º 109 (2020), p. 87-103. DOI: <https://doi.org/10.33426/rcg/2020/109/1438>
- Innerarity, Daniel. «The Data-Driven Pandemic: A new conceptualization of the Data Society». *STG Resilience Papers*, School of Transnational Governance (2021).

- Innerarity, Daniel. «Inteligencia artificial y democracia», en: Lazzarini, Carlos G. (dir.) *Inteligencia artificial y política. Los desafíos de una tecnología acelerada en las instituciones contemporáneas*. Buenos Aires: IcaP, 2022, p. 25-37.
- Innerarity, Daniel. «Justicia algorítmica y autodeterminación deliberativa». *ISEGORÍA, Revista de Filosofía moral y política*, n.º 68 (2023a), e23. DOI: <https://doi.org/10.3989/isegoria.2023.68.23>
- Innerarity, Daniel. «Predient el passat: una crítica filosòfica de l'anàlisi predictiva». *IDP, Revista de Internet, Derecho y Política*, n.º 39 (2023b), p. 1-11. DOI: <https://doi.org/10.7238/idp.v0i39.409672>
- Innerarity, Daniel. «The epistemic impossibility of an artificial intelligence takeover of democracy». *AI & Society*, vol. 39, (2023c), p. 1.667-1.671. DOI: <https://doi.org/10.1007/s00146-023-01632-1>
- Korinek, Anton y Stiglitz, Joseph E. «Artificial Intelligence and Its Implications for Income Distribution and Unemployment», en: Agrawal, Ajay; Gans, Joshua y Goldfarb, Avi (eds.) *The Economics of Artificial Intelligence: An Agenda*. Chicago: University of Chicago Press, 2019, p. 349-390.
- Lobato Rodríguez, Marta M. «Reseña de “El Capital en el Siglo XXI”, de Thomas Piketty». *Perifèria, Revista de Recerca i Formació en Antropologia*, vol. 19, n.º 2 (2014), p. 144-156.
- Maldita. «Qué datos personales puedes impedir que Facebook e IG utilicen para entrenar su inteligencia artificial (y cuáles no) y cómo hacerlo». *Maldita.es*, (6 de septiembre de 2023) (en línea) <https://maldita.es/malditatecnologia/20230906/datos-personales-meta-entrenar-IA/>
- Masut, Federica y Schmidt, Johanna (2024).« Inteligencia artificial, Responsabilidad humana. Documento de posición de Triodos Bank sobre la IA ética». *Triodos Bank*. [https://www.triodos.es/binaries/content/assets/tbes/vision-papers/24158\\_triodos\\_documento\\_vision\\_ia.pdf](https://www.triodos.es/binaries/content/assets/tbes/vision-papers/24158_triodos_documento_vision_ia.pdf)
- Mejías, Ulises A. y Couldry, Nick. «Datificación». *Revista Latinoamericana de Economía y Sociedad Digital*, (2022) (en línea). DOI: <https://doi.org/10.53857/hced6196>
- Monreal Garrido, Manuel. «Inteligencia Artificial. Implicaciones y aplicaciones en la Economía Social», en: Chaves Ávila, Rafael y Vañó Vañó, María José. *Dos decenios de actividad universitaria en economía social, cooperativismo y emprendimiento desde el instituto universitario IUDESCOOP*. Instituto Universitario de Economía Social, Cooperativismo y Emprendimiento (IUDESCOOP), CIRIEC-España, *Revista de Economía Pública, Social y Cooperativa* (2023). DOI: <https://doi.org/10.7203/10550/90499>
- Morozov, Evgeny. *Capitalismo Big Tech. ¿Welfare o neofeudalismo digital?* Madrid: Enclave de Libros, 2018.

- Moscardó Benavent, Susana. «Reseña de “La era del capitalismo de la vigilancia. La lucha de un futuro humano frente a las nuevas fronteras del poder”, de Shoshana Zuboff». *SCIO. Revista de Filosofía*, n.º 20 (2021), p. 265-269.
- Murrugarra Retamozo, Brenda Isabel. «Inteligencia artificial y privacidad en internet: amenazas para los datos personales de los usuarios». *Revista Científica Multidisciplinaria Ogma*, vol. 3, n.º 2 (2024), p. 30-48. DOI: <https://doi.org/10.69516/9dp8ap45>
- Naciones Unidas. «Los riesgos de la inteligencia artificial para la privacidad exigen medidas urgentes –Bachelet», Comunicado de prensa, (15 de septiembre de 2021) (en línea) <https://www.ohchr.org/es/press-releases/2021/09/artificial-intelligence-risks-privacy-demand-urgent-action-bachelet>
- Nikunen, Kaarina y Hokka, Jenni. «Welfare State Values and Public Service Media in the Era of Datafication». *Global Perspectives*, vol. 1, n.º 1 (2019) (en línea). DOI: <https://doi.org/10.1525/gp.2020.12906>
- Parlamento Europeo. «Artificial intelligence: threats and opportunities», (23 de septiembre de 2020) (en línea) <https://www.europarl.europa.eu/topics/en/article/20200918STO87404/artificial-intelligence-threats-and-opportunities>
- Pastor Sempere, Carmen. «La nueva Economía Social del Dato (ESD)». *CIRIEC-España, Revista Jurídica de Economía Social y Cooperativa*, n.º 41 (2022), p. 13-44. DOI: <https://doi.org/10.7203/CIRIEC-JUR.41.25571>
- Piketty, Thomas. *El Capital en el siglo XXI*. Madrid: FCE, 2014.
- Poell, Thomas; Nieborg, David B. y Van Dijck, José. «Platformisation». *Concepts of the Digital Society*, vol. 8, n.º 4 (2019) (en línea). DOI: <https://doi.org/10.14763/2019.4.1425>
- Prabhakaran, Vinodkumar; Mitchell, Margaret; Gebru, Timnit y Gabriel, Iason. «A Human Rights-Based Approach to Responsible AI». *ArXiv*, (2022), p. 1-17 (en línea) <https://arxiv.org/abs/2210.02667>
- Rhea, Alene; Markey, Kelsey; D'Arinzo, Lauren; Schellman, Hilke; Sloane, Mona; Squires, Paul y Stoyanovich, Julia. «Resume Format, LinkedIn URLs and Other Unexpected Influences on AI Personality Prediction in Hiring: Results of an Audit», en: *VVAA, Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, (2022), p. 572-587. DOI: <https://doi.org/10.1145/3514094.3534189>
- Sachs, Jeffrey D. *The Ages of Globalization: Geography, Technology and Institutions*. Nueva York: Columbia University Press, 2020.
- Salicrú Maltas, Joan. «Núria Oliver: Necessitem la IA per afrontar els grans reptes del segle XXI». *Revista Valors*, (7 de marzo de 2024) (en línea) <https://valors.org/necessitem-la-ia-per-afrontar-els-grans-reptes-del-segle-xxi/>

- Samoli, Sofia; López Cobo, Montserrat; Gómez Gutiérrez, Emilia; De Prato, Giuditta; Martínez-Plumed, Fernando y Delipetrev, Blagoj. *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*. Luxembourg; Publications Office of the European Union, 2020 (en línea) <https://publications.jrc.ec.europa.eu/repository/handle/JRC118163>
- Sassen, Saskia. *Expulsions. Brutality and Complexity in the Global Economy*. Cambridge, MA: Harvard University Press, 2014.
- Serrano Robles, Eloi. «Introducció a l'economia social», en: Serrano, Eloi (coord.) *Introducció a l'economia i l'empresa social*. Barcelona: Icaria, 2019, p. 1-.
- Serrano Robles, Eloi. *La contribución de la economía social a la sostenibilidad (una revisión metodológica para el análisis económico)*. Tesis doctoral, Universitat de les Illes Balears, 2024.
- Sharpa Marketing. «14 Herramientas de Inteligencia Artificial para LinkedIn». *Sharpa*, (2017) (en línea) <https://www.sharpamarketing.com/es/blog/14-herramientas-de-inteligencia-artificial-para-linkedin/>
- Starke, Christopher; Baleis, Janine; Keller, Birte y Marcinkowaki, Frank. «Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature». *Big Data & Society*, vol. 9, n.º 2 (2022), p. 1-16. DOI: <https://doi.org/10.1177/20539517221115189>
- Tolan, Songül. «Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges». *Digital Economy Working Paper* [Background paper to the European Commission's report: 'Artificial Intelligence: A European Perspective']. European Commission - Joint Research Centre (2019). <https://arxiv.org/abs/1901.04730>
- Toro, Oriol. «Núria Vallès: “Tot algoritme està vinculat a un sistema de valors”». *Revista Valors*, (1 de marzo de 2024) (en línea) <https://valors.org/tot-algoritme-esta-vinculat-a-un-sistema-de-valors/>
- Unión Europea. «Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial)». *Diario Oficial de la Unión Europea*, (12 de julio de 2024) (en línea) <http://data.europa.eu/eli/reg/2024/1689/oj>
- Van Dijck, José. «Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology». *Surveillance & Society*, vol. 12, n.º 2 (2014), p. 197-208. DOI: <https://doi.org/10.24908/ss.v12i2.4776>

- Varoufakis, Yanis. *Technofeudalism. What Killed Capitalism*. Londres: Vintage Books, 2023.
- Zajko, Mike. «Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates». *Sociology Compass*, vol. 16, n.º 3 (2022), e12962. DOI: <https://doi.org/10.1111/soc4.12962>
- Zhao, Jieyu; Wang, Tianlu; Yatskar, Mark; Ordoñez, Vicente y Chang, Kai-Wei. «Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints», en: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, (2017), p. 2.979-2.989. DOI: <https://doi.org/10.18653/v1/D17-1323>
- Zia, Tehseen. «Democratización de la IA: cómo impulsa a las personas y a las industrias [entrada de blog]». *Techopedia.*, (29 de diciembre de 2023) (en línea) <https://www.techopedia.com/es/democratizacion-ia>
- Zuboff, Shoshana. *La era del capitalismo de la vigilancia. La lucha por un futuro humano frente a las nuevas fronteras del poder*. Barcelona: Paidós, 2020.

# RELACIONES INTERNACIONALES

AÑO 33 - Nº 67 - Junio / Diciembre 2024

## DIÁLOGOS

*María de los Ángeles LASA*

## ESTUDIOS

*El fracaso de la política exterior populista: la política exterior brasileña en el gobierno de Jair Bolsonaro de 2019 a 2022*

**Beatriz Bandeira de Mello y Miriam Gomes Saraiva**

*Afganistán y los desafíos a la hegemonía estadounidense a finales del siglo XX y principios del XXI*

**Sidnei J. Munhoz**

*La insubordinación ideológica de la periferia. El caso de China*

**Juan Cruz Ramón Margueliche e Hilario Patronelli**

*Más calidad y menos cantidad en la Cooperación Internacional para el Desarrollo en el Sur Global: la experiencia de Brasil*

**Marina Bolfarine Caixeta y José Alejandro Sebastián Barrios Díaz**

*China como contrincante semi-periférico: recuperando el Estado en el análisis de sistemas-mundo*

**Luciano Moretti, Joel Sidler y Víctor Ramiro Fernández**

## DOSSIER

*Chile y Portugal en clave transnacional. 50 años del golpe de Estado y la Revolución de los claveles*

*Ecós del Cono Sur sobre la Revolución de los Claveles (1974-1976): La experiencia chilena y el velascato peruano*

**Ricardo Andrés Pérez Haristoy y Gilberto Aranda Bustamante**

*La trascendencia de la guerra colonial y el golpe de Estado del 25 de abril de 1974 en las relaciones entre Cuba y Portugal (1959-1975)*

**Fernando Camacho Padilla**

*De septiembre a abril. El golpe militar de Chile y la Revolución de los claveles portuguesa bajo la mirada de La Voz de Galicia*

**Francisco Javier Morales Aguilera**

*La izquierda española frente al golpe de Estado en Chile y la Revolución de los Claveles: el caso de los marcos interpretativos de LC y ORT*

**Miguel Alejandro Pérez Cabrera**

## REFLEXIONES

*Futuro incierto: regresar al medioevo o crear algo nuevo*

**Javier L. Surasky**

Facultad de Ciencias Jurídicas y Sociales

48 n° 582 5° piso (1900)  
La Plata REPUBLICA ARGENTINA  
Tel 54 221 6442096



Instituto de  
Relaciones  
Internacionales



UNIVERSIDAD  
NACIONAL  
DE LA PLATA

# La regulación europea de la IA ante los sesgos y riesgos de discriminación algorítmica en contextos migratorios

## The European regulation on AI in the face of the biases and risks of algorithmic discrimination in migration contexts

**Encarnación La Spina**

Profesora contratada doctora encargada, Facultad de derecho, Universidad de Deusto (Bilbao).  
elaspina@deusto.es. ORCID: <https://orcid.org/0000-0003-0423-0674>

**Cómo citar este artículo:** La Spina, Encarnación. «La regulación europea de la IA ante los sesgos y riesgos de discriminación algorítmica en contextos migratorios». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 171-194. DOI: doi.org/10.24241/rcai.2024.138.3.171

**Resumen:** La creciente institucionalización de la inteligencia artificial (IA) es un problema social especialmente preocupante para aquellos colectivos más vulnerables, como pueden ser las personas migrantes y solicitantes de asilo. Si bien los contextos migratorios globales están siendo objeto de un férreo control securitario y de una amplia transformación digital, la regulación europea sobre la IA presenta todavía un alcance muy limitado para contrarrestar las repercusiones negativas del diseño y entrenamiento de algoritmos. A tal propósito, este artículo plantea una revisión crítica sobre cómo se introducen los sistemas de IA en el control y la gestión migratorios europeos y analiza por qué el nuevo marco regulador de la IA es insuficiente para evitar el impacto discriminatorio de sesgos, estereotipos y riesgos asociados a las dinámicas asimétricas de poder y a las estructuras de injusticia social subyacentes a las migraciones.

**Palabras clave:** migraciones, IA, derechos humanos, discriminación algorítmica, sesgos, estereotipos, Unión Europea (UE)

**Abstract:** *The increasing institutionalisation of artificial intelligence (AI) is a social problem of particular concern for the most vulnerable groups, such as migrants and asylum seekers. While global migration contexts face tight security control and an extensive digital transition, the European regulation on AI is still too limited in scope to counteract the negative impacts of the design and training of algorithms. With that in mind, this paper provides a critical review of how AI-driven systems are introduced into European migration control and management, and analyses why the new legal framework for AI does not do enough to prevent the discriminatory impact of biases, stereotypes and risks associated with the asymmetric power dynamics and the structures of social injustice underlying migration.*

**Key words:** *migrations, AI, human rights, algorithmic discrimination, bias, stereotypes, European Union (EU)*

*Este trabajo se ha realizado con el apoyo del Proyecto Derechos Humanos y retos socioculturales en un mundo en transformación (ref.: IT1468-22) y del Proyecto I+D+i RESEST «Resiliencia del derecho antidiscriminatorio a los sesgos y estereotipos: desafíos y propuestas de intervención» (ref. PID2021-123171OB-I00), del Ministerio de Ciencia e Innovación.*

Un primer abordaje sobre el trabajo de la Unión Europea (UE) referente a la regulación de la inteligencia artificial (IA), requiere entender qué es la IA y qué comprende un sistema de IA; sin embargo, no existe un consenso general sobre su definición (Ruiz Tarrías, 2023: 95). Entre las varias descripciones técnicas posibles, aquellas más extendidas, si bien con significaciones muy diferentes, sostienen que la IA se refiere a un conjunto de tecnologías que utilizan datos combinados y aplican modelos o algoritmos, junto con el poder de cómputo proporcionado por las Tecnologías de la Información (TI), para procesar y generar información de manera eficiente a partir de gran cantidad de datos de entrada o salida. Su finalidad principal es obtener con cierto grado de autonomía y precisión, resultados sólidos, como son predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales más o menos concretos que requerirían, por lo general, la inteligencia humana para completarse satisfactoriamente (Zuiderveen Borgesius, 2018; FRA, 2022). Precisamente, dada la capacidad de inferencia y el alcance de la IA frente a *softwares*, *hardwares* o programaciones tradicionales, la UE trabaja desde hace años en una regulación de la IA que sea pionera a escala global, multisectorial y transversal, tanto en el plano institucional como social (Ruiz Tarrías, 2023: 92).

Uno de los ámbitos de acción prioritaria de la agenda europea con potencial impacto de la IA es la gestión de los flujos migratorios de llegada, el control y la externalización de las fronteras en países de origen y tránsito, la verificación de la identidad, así como la atribución del estatuto de residencia, sin perjuicio de los procesos de determinación del asilo o de acogida e integración. En este amplio ámbito competencial, el uso de tecnologías basadas en sistemas de IA –una de las herramientas de transformación digital más potente y de rápida evolución– ya está prácticamente generalizado y casi plenamente operativo a escala europea (EMN, 2022: 11; Okzul, 2023: 64 y 40); al igual que en Canadá, Australia, China y Estados Unidos (Akhmetova, 2020; Nalbandian, 2022; Molnar, 2024). En la UE, el uso progresivo de la IA es ya una realidad en estados miembros pioneros como Alemania, Finlandia, Hungría, Grecia, Letonia, Lituania y los Países Bajos<sup>1</sup>, pero también en diferentes agencias europeas como la Agencia para la Gestión Operativa de Sistemas Informáticos de Gran Magnitud en el Espacio de Libertad Seguridad y Justicia (eu-LISA),

---

1. Según EMN (2022: 10-11), se aplican los siguientes sistemas de reconocimiento del habla: DIAS en Alemania; TILDE en Letonia; sistemas de identificación de identidad TIKKA en Finlandia; y reconocimiento facial en Grecia, Hungría y Letonia.

la Agencia de Asilo de la UE (AAUE), la Agencia de la Guardia de Fronteras y Costas de la UE (Frontex), así como en otras organizaciones internacionales, como ACNUR, el Comité de Rescate Internacional (IRC, por sus siglas en inglés) y la Organización Internacional para las Migraciones (OIM), que se sirven de la IA para gestionar una gobernanza migratoria más eficiente, neutral y objetiva (Marin, 2011: 145; Leese, 2014: 495).

Ahora bien, este avance exponencial de la IA no sintoniza con el ritmo de transformación lento que marca el Reglamento de IA de la UE del 13 de junio de 2024<sup>2</sup>. Es más, con su escalonada entrada en vigor a partir de 2026, se confirma como objetivo prioritario apostar por un uso fiable de la IA y, en menor medida, prohibir todos los posibles impactos negativos asociados a la aplicación de estos sistemas de *software* inteligente. Por ello, aunque la entidad de los múltiples riesgos asociados a la discriminación algorítmica es ya de sobra conocida y advertida (Chander, 2017), su presunta inocuidad, exactitud u objetividad son cada vez más cuestionadas por la doctrina. Por ejemplo, por quienes sostienen su posible incompatibilidad *prima facie* en términos de protección de datos y privacidad y, en segunda instancia, por los que alertan de su inevitable impacto discriminatorio en función de sus posibles fines y datos, personas o grupos afectados y contextos de aplicación. Precisamente, el difícil equilibrio entre el carácter no discriminatorio y la transparencia de los sistemas de IA (EDRi, 2022) vaticina una futura tendencia regresiva y menos garantista hacia la protección de los derechos humanos en la UE, especialmente para ciertas personas o grupos vulnerables (Avello, 2023; Garrido Carillo, 2023).

Por esta razón, se ha empezado a estudiar no solo cómo se diseñan, desarrollan e implementan los diferentes sistemas de IA de externalización y control fronterizo (Leese *et al.*, 2022; Napolitano, 2023: 7-8) —aquellos usados en ciclos migratorios de tránsito y destino para la acogida extraordinaria u ordinaria (Beduschi, 2022; Leese, 2014)—, sino que también se ha puesto el foco sobre los desafíos ético-jurídicos inherentes a la IA (FRA, 2022; Moreno, 2021; Cruz, 2020 Vavoula,

**En la UE, el uso progresivo de la IA es ya una realidad en estados miembros pioneros como Alemania, Finlandia, Hungría, Grecia, Letonia, Lituania y los Países Bajos, pero también en diferentes agencias europeas.**

---

2. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial. Véase: [https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ%3AL\\_202401689](https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ%3AL_202401689)

2021). En particular, estos desafíos son evidentes en los contextos fronterizos y en la gestión migratoria de entrada/salida/estancia (Beduschi, 2021; Molnar, 2023), al convertirse en el mayor laboratorio de datos a gran escala que existe para el diseño de *software* algorítmico y para la progresiva experimentación de sistemas de IA (EC, 2011, 2016, 2021a, 2021b y 2021c). Y ello porque, básicamente, estos sistemas parecen abocados a tolerar sine die dinámicas asimétricas de poder sobre los cuerpos migrantes, lo que ya ha demostrado ser altamente discriminatorio y sesgado, sobre todo contra grupos racializados y marginados sin libertad de decisión o autonomía migratoria, por no mencionar que aquellos son culturalmente insensibles a la diversidad (Molnar, 2019; Forti, 2024; Romano, 2023).

A fin de analizar este fenómeno, este artículo plantea, en primer lugar, una retrospectiva que examina cómo se ha producido la transformación digital en los sistemas de información europeos a gran escala que operan en el ámbito migratorio, incluyendo especialmente un mapeo de los proyectos financiados por la UE que ya incluyen tecnología IA en ciertos países; y, en segundo lugar, al hilo del Reglamento de IA de la UE, se identifica y se valora cómo se regulan en este nuevo marco normativo los factores de discriminación algorítmica y cómo se controlan los sesgos asociados al uso de tales sistemas de IA en contextos migratorios.

## La transformación digital y el uso de la IA en el ámbito migratorio europeo

Desde la puesta en marcha del Espacio de Libertad, Seguridad y Justicia (ELSJ; Tratado de Ámsterdam, 1999), la UE ha mantenido, pese a un desarrollo normativo y programático cambiante, una línea continuista de carácter securitario sobre la migración y el asilo. Una muestra de esta creciente deriva y apuesta por la seguridad ha sido la paulatina aplicación de diferentes dispositivos tecnológicos y biométricos basados en sistemas de IA capaces de mejorar el monitoreo, así como la eficiencia de los procesos y las prácticas de securitización, externalización y criminalización en la UE (art. 3.2 TUE y art. 67 y 77 TFUE). De este modo, el ELSJ, por medio de la transformación digital de estas prácticas, especialmente en las fronteras exteriores, se ha convertido en un complejo ecosistema tecnológico, digital e inteligente gracias al desarrollo actual de la gestión integrada y a la interoperabilidad de los diferentes sistemas de información a gran escala (Broeders, 2007; Vavoula, 2021), así como al progresivo uso aplicado de la IA en el marco de proyectos financiados con fondos europeos.

## Principales hitos en la transformación digital de la gestión integrada de los sistemas de información a gran escala

El objetivo de reforzar los sistemas de información, la arquitectura de datos y el intercambio de información en los ámbitos de la gestión de las fronteras y la lucha contra el terrorismo ha sido determinante para consolidar el proceso de transformación digital europeo. Sin duda, un primer hito de este proceso se materializó en 2022 con la última actualización del Sistema de Información Schengen (SIS), con el SIS II<sup>3</sup> que, pese a limitar el acceso a los datos estrictamente necesarios para determinadas autoridades en cumplimiento de su misión (Cruz Ángeles, 2022: 94-104), incluye el acceso a datos biométricos (huellas palmares, imágenes faciales y perfiles de ADN relacionados con personas desaparecidas o con infracciones penales) a efectos de denegación de entrada o de estancia en el espacio Schengen.

En paralelo, un segundo hito ha sido la reciente reforma de la base de datos de dactiloscopia de solicitantes de asilo (EURODAC)<sup>4</sup>, un sistema de determinación del país responsable del examen de las solicitudes de protección internacional, pero que también permite el acceso policial para prevenir, detectar e investigar delitos de terrorismo y otros delitos graves. Con su entrada en vigor en 2026, será posible incluir la recogida de datos alfanuméricos de identidad que permitan el recuento de personas solicitantes de asilo (no solo de solicitudes); la inclusión de nuevas categorías de personas (aquellas desembarcadas tras operaciones de búsqueda y rescate para un control preliminar en frontera); el tratamiento de nuevos datos biométricos (imagen facial y ADN); así como la reducción de la edad para la toma de huellas dactilares (de los 12 a los 6 años).

Como tercer hito más próximo a un proceso avanzado de *securitización* tecnológica (Oliveira y Gabrielsen, 2022: 1443), cabe mencionar la reciente reforma del Sistema de Información de Visados (SIV)<sup>5</sup>. En concreto, esta reforma ha permitido ampliar su ámbito de aplicación (incluir datos sobre visados de larga duración y permisos de residencia) y mejorar los controles en la tramitación de visados (entre otros, con la comprobación de antecedentes

---

3. Reglamento (UE) 2022/1190 del Parlamento Europeo y del Consejo de 6 de julio de 2022 (...) *DO L* 185, del 12 de julio de 2022.

4. Reglamento UE 2024/1358 del Parlamento Europeo y del Consejo, de 14 de mayo de 2024 (...) *DO L*, 2024/1350, del 22 de mayo de 2024.

5. Reglamento (UE) 2021/1133 del Parlamento Europeo y del Consejo de 7 de julio de 2021 (...) *DO L* 248, del 13 de julio de 2021.

para identificar riesgos), dado que, en determinadas condiciones, las autoridades nacionales y Europol pueden acceder a los datos del SIV con fines de prevención, detección e investigación de delitos terroristas y penales.

Asimismo, desde 2019 y dentro del proceso de transformación digital, cabe señalar un cuarto hito con la creación de otros tres nuevos sistemas (EES, ETIAS y ECRIS-TCN, por sus siglas en inglés). El primero, el Sistema de Entradas y Salidas (EES)<sup>6</sup>, fue establecido en 2017 para registrar digitalmente la entrada y salida (y la denegación de entrada) de los titulares de visados de corta duración y de los viajeros exentos de visado que crucen las fronteras exteriores de la UE. Este sistema calcula automáticamente la duración de la estancia autorizada de una persona y genera alertas a los estados miembros cuando la estancia autorizada haya expirado; además, las autoridades policiales y Europol podrán acceder al EES para prevenir, detectar e investigar delitos de terrorismo u otros delitos graves. El segundo, el Sistema Europeo de Información y Autorización de Viajes (ETIAS)<sup>7</sup>, implementado en 2018, preregistra a los visitantes exentos de visado que viajan al espacio Schengen y así evalúa los riesgos de seguridad o de migración irregular que plantean estas personas antes de que lleguen a la frontera por medio de una lista de vigilancia de personas sospechosas. Y, el tercero, el Sistema Europeo de Información de Antecedentes Penales de Nacionales de Terceros Países (ECRIS-TCN) creado en 2019<sup>8</sup>, permite saber qué otros estados miembros poseen antecedentes penales de nacionales de terceros países o apátridas, o de ciudadanos de la UE que también posean la nacionalidad de un tercer país.

Por último, como quinto hito, se añade el desarrollo y la revisión de un marco de interoperabilidad (EC, 2016) regulado en el ámbito de los Reglamentos 2019/817 y 2019/818<sup>9</sup> que incluye la puesta en marcha de cuatro componentes

- 
6. Reglamento (UE) 2017/2226 del Parlamento Europeo y del Consejo, de 30 de noviembre de 2017 (...) DO L 327, del 9 de diciembre de 2017 modificado por el Reglamento (UE) 2024/1356 del Parlamento y del Consejo de 14 de mayo de 2024 (...) DO L, 2024/1350, del 22 de mayo de 2024.
  7. Reglamento (UE) 2018/1240 del Parlamento Europeo y del Consejo, de 12 de septiembre de 2018 (...) DO L 249, del 14 de julio de 2021, modificado por el Reglamento (UE) 2024/1356 del Parlamento y del Consejo de 14 de mayo de 2024 (...) DO L, 2024/1350, del 22 de mayo de 2024.
  8. Reglamento (UE) 2019/816 del Parlamento Europeo y del Consejo, de 17 de abril de 2019 (...) DO L 135, del 22 de mayo de 2019 modificado por el Reglamento (UE) 2024/1352 del Parlamento y del Consejo de 14 de mayo de 2024 (...) DO L, 2024/1350, del 22 de mayo de 2024.
  9. El primero ha sido modificado por el Reglamento (UE) 2021/1150 del Parlamento Europeo y del Consejo del 7 de julio de 2021 (...) DO L 249, del 14 de julio de 2021. Mientras que, el segundo, por el Reglamento (UE) 2021/1151 del Parlamento Europeo y del Consejo de 7 de julio de 2021, DO L 249, del 14 de julio de 2021, el Reglamento (UE) 2021/1152 del Parlamento Europeo y del Consejo de 7 de julio de 2021 (...) DO L 249, del 14 de julio de 2021 y el Reglamento (UE) 2024/1352 del Parlamento y del Consejo de 14 de mayo de 2024 (...) DO L, 2024/1350, del 22 de mayo de 2024.

cruzados: el Portal Europeo de Búsqueda (PES, por sus siglas en inglés), el Servicio Compartido de Correspondencias Biométricas (BMS, por sus siglas en inglés), el Repositorio Común de Identidades (CIR, por sus siglas en inglés) y el Detector de Identidades Múltiples (MID, por sus siglas en inglés) (Dumbrava, 2021: 17; Cruz Ángeles, 2022).

## **La aplicación de la IA en los proyectos europeos financiados para la gestión y control de los flujos migratorios**

Una vez alcanzada la plena interoperabilidad de los mencionados sistemas en 2025, y pese a las críticas vertidas sobre el proceso algorítmico del ETIAS (Guild y Vavoula, 2020; Derave *et al.*, 2022), la irrupción de sistemas de IA es cada vez más tangible en la UE. Precisamente, la financiación pública que reciben numerosos proyectos a escala europea (Kilpatrick y Jones, 2022: 31-46) ha hecho posible incorporar técnicas de aprendizaje-máquina que generan contenidos, predicciones, recomendaciones o decisiones con capacidad de inferencia en el entorno con el que el sistema interactúa. Por ejemplo, a día de hoy, ya se ha testado y desarrollado IA con *software* algorítmico aplicable al control de flujos migratorios, la evaluación automatizada de riesgos, la predicción por medio de polígrafos y vigilancia por drones, cámaras termográficas, sensores de movimiento y térmicos o bien los sistemas de identificación y categorización biométrica en remoto, la captación y reconocimiento de imágenes faciales, emociones, idiomas o la recopilación automática de huellas dactilares u otros datos personales (Burgess y Klotza, 2021; Musco, 2023: 25).

**Ya se ha testado y desarrollado IA con *software* algorítmico aplicable al control de flujos migratorios, la evaluación automatizada de riesgos y la predicción, con sistemas de identificación y categorización biométrica en remoto, la captación y reconocimiento de imágenes faciales, emociones, idiomas o la recopilación automática de huellas dactilares.**

Tal y como se recoge en la tabla 1, no solo hay variedad de técnicas de IA en función del grado de eficiencia de los objetivos perseguidos (prever la movilidad humana o predecir la demanda futura en los sistemas nacionales de asilo, informar de las operaciones de control fronterizo o coordinar ayuda humanitaria), sino que también se puede observar el cumplimiento de diferentes funciones (Loddo y Addis, 2023). Entre otras, destacan las funciones predictivas o constrictivas y preventivas en el control fronterizo frente a las funciones informativas o procedimentales en la gestión migratoria.

Tabla 1. Proyectos europeos que usan IA

Objetivos	Función	Países/ciudades	Técnicas
TRESSPASS (2018-2021) <a href="https://cordis.europa.eu/project/id/787120">https://cordis.europa.eu/project/id/787120</a>	Vigilancia fronteriza; lucha contra delitos fronterizos y otras amenazas	Reino Unido	Sistema digital integrado de gestión de frontera basado en un software algorítmico para detectar el riesgo.
CENTAUR - HYPERION (2020-X)*	Vigilancia fronteriza	Grecia, (Lesbos, Chios, Samos, Leros y Kos)	Sistema digital integrado de gestión de la seguridad electrónica y física colocado dentro y alrededor de las instalaciones (campos de refugiados) mediante cámaras y un algoritmo de análisis del movimiento (AI Behavioral Analytics). <ul style="list-style-type: none"> <li>- 112 cámaras</li> <li>- 112 altavoces</li> <li>- 19 escáneres de rayos X</li> <li>- 55 puertas magnéticas</li> <li>- 20 cámaras con análisis de comportamiento</li> <li>- 94 cámaras para entradas y salidas</li> <li>- 7 drones</li> <li>- Análisis de datos biométricos</li> </ul>
SOLOMON (2017-2019) <a href="https://cordis.europa.eu/project/id/705020">https://cordis.europa.eu/project/id/705020</a>	Vigilancia fronteriza	Reino Unido	Red de cámaras inteligentes como el VideoBadge de Edesix, se combinan con las capacidades de procesamiento de los teléfonos inteligentes. Estos dispositivos los llevan policías o guardias de seguridad o se montan en robots móviles.
REBORDER (2017-2021) <a href="https://www.reborder.eu/">https://www.reborder.eu/</a>	Vigilancia fronteriza	Grecia	Robots móviles sin equipaje, y se incluyen vehículos aéreos, acústicos, subacuáticos y terrestres, en grado de operar tanto autónomamente, incorporando sensores multimodales, como parte de una red interoperacional.
ARESIBO (2019-2022) <a href="https://cordis.europa.eu/project/id/833805">https://cordis.europa.eu/project/id/833805</a>	Análisis predictivo de movimientos	Finlandia, Grecia, Rumania, Portugal	Sistemas no tripulados (inteligencia de enjambre UxV) que pueden satisfacer los requisitos operativos de diferentes misiones de seguridad fronteriza en las fronteras verdes y azules.
iBORDERCTRL (2016-2019) <a href="https://www.iborderctrl.eu">https://www.iborderctrl.eu</a>	Reconocimiento de emociones	Hungría, Grecia, Letonia	Sistema automatizado de control fronterizo con avatares detectores de mentiras. Nuevas tecnologías biométricas avanzadas: cotejo facial y escaneo de las venas de la palma de la mano.
FOLDOUT (2018-2022) <a href="https://cordis.europa.eu/project/id/787021">https://cordis.europa.eu/project/id/787021</a>	Análisis predictivo de movimientos	Bulgaria, Finlandia, Grecia, Guyana F.	Sistema de fusión y seguimiento de múltiples sensores que utiliza detectores infrarrojos pasivos en combinación con la detección automática de personas a partir de imágenes de cámaras de video térmicas y visuales.

STARLIGHT (2021-2025) <a href="https://cordis.europa.eu/project/id/101021797">https://cordis.europa.eu/project/id/101021797</a>	Gestión de la seguridad, incluido el ámbito de las migraciones	Procedimental	España, Austria, Francia, Alemania, Países Bajos, Polonia, Bélgica, Lituania, Finlandia, Suecia, Estonia, Letonia, Grecia, Chipre, Portugal	Herramientas de IA para las fuerzas y cuerpos de seguridad que facilite la comprensión, las herramientas, los datos y la protección adecuados (AI-based solutions).
NADINE (2018-2021) <a href="https://nadine-project.eu">https://nadine-project.eu</a>	Apoyo integración	Informativa	Francia, Italia, Reino Unido, Grecia, Bélgica, Luxemburgo, España	Herramientas TIC para la recogida de datos sobre las personas migrantes y refugiadas relacionadas con las competencias y aptitudes, junto con otra información que afectará a su procedimiento de gestión de la integración.
REBUILD (2019-2022) <a href="https://www.rebuild.europa.eu/en/default.aspx">https://www.rebuild.europa.eu/en/default.aspx</a>	Apoyo integración	Informativa	Italia, España, Grecia	Herramientas TIC que proporcionen apoyo tanto a las personas migrantes y refugiadas en Europa para acceder a los servicios existentes, como a los proveedores de servicios locales que interactúan con personas migrantes y refugiadas para mejorar su prestación de servicios.
MIICT (2018-2022) <a href="https://www.miict.eu/project-info">https://www.miict.eu/project-info</a>	Apoyo integración	Informativa	Reino Unido	Servicios mejorados basados en las TIC con personas migrantes, refugiadas, solicitantes de asilo y servicios del sector público.
MICADO (2019-2022) <a href="https://www.micadoproject.eu/">https://www.micadoproject.eu/</a>	Apoyo integración	Informativa	Alemania, Polonia, Hungría, República Checa, Eslovenia, Austria, Italia, Francia, España, Reino Unido	Herramientas que aborden el reto de la integración de las personas migrantes mediante la creación conjunta de servicios mejorados basados en las TIC con personas migrantes, refugiadas, solicitantes de asilo y servicios del sector público
WELCOME (2020-2023) <a href="https://www.welcome-h2020.eu/">https://www.welcome-h2020.eu/</a>	Apoyo integración	Informativa	España, Alemania, Grecia	Servicios múltiples de agentes de conversación inteligentes para la acogida, gestión e integración de nacionales de terceros países en la UE.
EASYRIGHTS (2019-2022) <a href="https://www.easyrights.eu">https://www.easyrights.eu</a>	Apoyo integración	Informativa	Austria, Dinamarca, Grecia, Italia, Noruega, España, Reino Unido	Servicios personalizados de reconocimiento de voz y NIPT (Natural language Processing tools).
KRISTINA (2015-2018) <a href="https://cordis.europa.eu/project/id/645012/es">https://cordis.europa.eu/project/id/645012/es</a>	Apoyo integración	Informativa	España, Alemania	Tecnologías de comunicación móviles que sirven a las personas migrantes con barreras lingüísticas y culturales en el país de acogida en cuestiones relacionadas con la atención básica y la asistencia sanitaria.

\* No hay información en el portal CORDIS de la UE. Ambos sistemas se desplegaron sin el nombramiento previo de un responsable de la protección de datos ni una evaluación obligatoria del impacto. Además, las autoridades de control (EDPB-EDPS) declararon en el Dictamen conjunto 5/2021 que las tecnologías de reconocimiento del comportamiento deberían prohibirse debido al riesgo inaceptable que suponen para los derechos fundamentales. Véase la pregunta presentada con solicitud de respuesta escrita E-003094/2022 a la Comisión europea: [https://www.europarl.europa.eu/doceo/document/E-9-2022-003094\\_EN.html#def1](https://www.europarl.europa.eu/doceo/document/E-9-2022-003094_EN.html#def1) [fecha de consulta: 26.07.2024].  
Fuente: elaboración propia a partir de datos del sitio web de EU research CORDIS.

Por último, también existen distintos informes (EMN, 2022; EC, 2020a y 2021b) que recogen evidencias de otros usos aplicados de la IA desde administraciones públicas y ONG. Por ejemplo, *chatbots*<sup>10</sup>, asistentes virtuales o interactivos de voz, *software* algorítmico para la logística de centros de acogida o programas de integración<sup>11</sup>. Todos ellos, conforme al marco normativo vigente, deberían cumplir con salvaguardias éticas, obligaciones de transparencia o reducción de riesgos dado el volumen de características y datos personales (edad, sexo, tamaño de la familia, idioma, vulnerabilidades) que almacenan, manejan o acceden antes del proceso de entrada y durante la estancia en la UE.

## El Reglamento de IA de la UE ante el uso de la IA en el contexto migratorio europeo

Frente a la funcionalidad y versatilidad de los mencionados sistemas basados en IA en el ámbito migratorio, los modelos regulatorios de la IA han quedado absortos por una amplia e infructuosa discusión sobre tres diferentes opciones: autorregulación, hiperregulación y *laissez faire* (Ruiz Tarrías, 2023: 93). Siguiendo un enfoque equilibrado entre derechos y beneficios para una plena «soberanía europea digital», la legislación europea en materia de IA ha sido sometida a una ardua negociación para construir y permitir de facto un modelo de vigilancia regulatoria casi abstencionista (ibídem: 115), habida cuenta de las limitadas prohibiciones y las excepciones vigentes en contextos migratorios (EDRI, 2022). Así pues, actualmente, el Reglamento de IA de la UE configura como techo de cristal un «sistema de regulación de cinco niveles basado en el riesgo», que abarca desde la ausencia de regulación en el caso de los sistemas de IA más inocuos hasta la prohibición absoluta en el caso de los más peligrosos como, entre otros, las captaciones de rasgos de identidad automatizadas (art. 5.1).

---

10. *Chatbots* de asistente virtual como NATALI y LEPINE en Francia; DEMS en los Países Bajos; MIGAPP de la OIM; REFUGEE BUDDY de la Cruz Roja en los Países Bajos, Noruega y Canadá; LOVE EUROPE; TIMEPEACE APP en el Reino Unido; ANKOMMEN en Alemania; FINDAWAY en el sur de Francia; MOBILEARN en Suecia; KAMU en Finlandia; (RE)settle in Croatia, en Croacia (EMN, 2022).

11. Entre otros, MATCH'IN en Alemania; GEOMATCH en Suiza y los Países Bajos, y MOT en Noruega (Okzul, 2023).

Sin embargo, tal y como advierten las voces más críticas (Veale y Zuiderveen Borgesius, 2021), para grupos históricamente marginados y vulnerables como las personas migrantes es necesario calibrar el impacto discriminatorio de los algoritmos aplicados a tales sistemas, analizando tanto el anverso de este marco regulador como sus posibles reversos.

## **El anverso limitado del Reglamento de IA en materia de migración, asilo y control fronterizo**

El Reglamento de IA inscribe la migración, el asilo y el control fronterizo dentro de las áreas de alto riesgo para los derechos fundamentales, y recoge obligaciones de transparencia y comunicación de información a los usuarios (art. 13), así como la necesidad de vigilancia humana (art. 14). Sin embargo, de un lado, limita su ámbito de aplicación a las personas migrantes afectadas que estén ubicadas en la UE (art. 2.1 g), es decir, obvia las «ficciones de no entrada» previstas en el Reglamento de triaje (*screening*)<sup>12</sup>, y excluye los proyectos de investigación y desarrollo científico (véase tabla 1) que no se hayan introducido al mercado (art. 2.6); y, de otro lado, no contempla por ahora la prohibición absoluta de evaluación de indicadores de riesgos, ni tampoco se prevé la necesidad de auditorías externas de sistemas de IA para poder detectar y corregir posibles sesgos discriminatorios «de origen racial o étnico» (UN, 2020; EDRI, 2022). Esta aseveración se deriva del artículo 5 donde se recoge un elenco de prácticas prohibidas (total o parcialmente matizadas) y riesgos inaceptables, pero al mismo tiempo se mantienen vía el artículo 6.2 amplias excepciones respecto a la necesidad de autorización previa (art. 5.3), el requisito de doble verificación (art. 14.5) y la confidencialidad de la información intercambiada (art. 58.3; art. 59.1 f), que reducen a mínimos el alcance de lo prohibido frente a lo tolerado y permitido en los sistemas de IA de alto riesgo del Anexo III apartado 7.

**El Reglamento de IA no contempla por ahora la prohibición absoluta de evaluación de indicadores de riesgos, ni tampoco se prevé la necesidad de auditorías externas de sistemas de IA para poder detectar y corregir posibles sesgos discriminatorios «de origen racial o étnico».**

---

12. Reglamento (UE) 2024/1356 del Parlamento Europeo y del Consejo, del 14 de mayo de 2024, por el que se introduce el triaje de nacionales de terceros países en las fronteras exteriores y se modifican los Reglamentos (CE) n.º. 767/2008, (UE) 2017/2226, (UE) 2018/1240 y (UE) 2019/817.

En concreto, figurarían como prohibidos (art. 5.1), en primer lugar, los sistemas de IA que se sirvan de técnicas subliminales que trasciendan la conciencia humana para alterar de manera sustancial su comportamiento de un modo que provoque incluso de forma probable perjuicios físicos o psicológicos a esa persona u otra. En segundo lugar, sistemas de IA que aprovechen alguna de las vulnerabilidades de un grupo específico de personas debido a su edad o discapacidad física o mental para alterar de manera sustancial el comportamiento de una persona que pertenezca a dicho grupo de un modo que provoque incluso de forma probable perjuicios físicos o psicológicos a esa persona u otra. En tercer lugar, sistemas de IA usados por parte de las autoridades públicas con el fin de evaluar o clasificar la fiabilidad de personas físicas durante un período determinado de tiempo atendiendo a su conducta social o económica, a características personales o de su personalidad conocidas que provoquen: un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos enteros y que este sea injustificado o desproporcionado. Y, por último, la utilización de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines policiales, salvo que sea estrictamente necesario para búsqueda de posibles víctimas, prevención de amenazas a la seguridad, o búsqueda de posible perpetrador o sospechoso.

Si se hace una triangulación entre las mencionadas herramientas IA recogidas en los proyectos ya implementados en la UE, todas ellas integrables en el Anexo III 1, 6 y 7, y este catálogo general de prohibiciones, se erigen importantes salvedades, gracias a la eficacia de los usos y fines de aquellas prácticas que devienen legales conforme con el derecho nacional y de la UE (considerando 30 y art. 6.3). Por ejemplo, sería el caso de los proyectos que cumplen una función predictiva como iBORDERCTRL, para detectar la credibilidad del estado emocional de una persona física, o proyectos que se sirven de herramientas predictivas (FOLDOUT o ARESIBO); así como la gran mayoría de aquellos que persiguen la función preventiva<sup>13</sup> para evaluar determinados riesgos que presenten personas físicas que entren en el territorio de un Estado miembro o soliciten un visado o asilo; sin perjuicio de los desarrolladores algorítmicos previstos en los sistemas de información a gran escala (Anexo X), que permiten verificar la autenticidad de los documentos pertinentes de personas físicas o bien ayudan a las autoridades públicas competentes a examinar las solicitudes de asilo, visado y permiso de residencia, así como las reclamaciones conexas para determinar si las personas físicas solicitantes de asilo reúnen los requisitos necesarios para su obtención.

De este modo, la valencia de las prohibiciones y los usos admisibles parecen entrar en neta contradicción con la letra de los considerandos 6, 7 y 60 del Reglamento de

---

13. Entre otros, véase el proyecto Tresspass, Centaur, Hyperion, Solomon.

IA de la UE, especialmente con su referencia a los valores y derechos europeos y los principios *ius cogens* de la protección del asilo. Básicamente porque, primero, desde la excepcionalidad se reconoce que tales sistemas mayoritarios, aun teniendo un vago soporte garantista (Romano, 2023: 240), sirven «para preservar la capacidad de las autoridades competentes (...)» (considerando 33), pero son y pueden ser invasivos al afectar significativamente la vida de personas migrantes que a menudo se encuentran en una posición especialmente vulnerable (considerando 30) o generar formas de discriminación no solo de origen racial o étnico (considerando 56). Y, segundo, porque las prohibiciones de alguno de ellos pueden modularse o justificarse por razones de eficiencia y oportunidad como «aliviar la carga administrativa y no retrasar la implementación de sistemas del Anexo x». Este es un *desiderátum* vacío sobre la exactitud, el carácter no discriminatorio y la transparencia de los sistemas de IA utilizados, que atribuye a quienes diseñan y controlan los sistemas de IA, una fiabilidad y autocontención técnica para garantizar el respeto de los derechos fundamentales de las personas o grupos vulnerables afectados, sin reforzar los mínimos exigibles en caso de advertirse un impacto discriminatorio directo o indirecto.

Por último, resulta contradictorio con la voluntad de impulsar «una tecnología centrada en el ser humano» que los sistemas de información ya introducidos en el mercado o puestos en servicio según el Anexo x (SIV, SIS, Eurodac, SES, SEIAV o ETIAS), queden excluidos de vigilancia regulatoria según el art. 11.1 «hasta el 31 de diciembre de 2030». Máxime si el art. 6.3 con su referencia al Anexo III apartado 7 reconoce como objetivo y razonable que las autoridades encargadas del control fronterizo, de la migración y del asilo, sí pueden «proceder a evaluar riesgos de migración irregular, seguridad y salud». Estas autoridades encargadas de la aplicación de la ley «deben poder utilizar sistemas de información, de conformidad con el derecho de la Unión o el derecho nacional, para identificar a una persona que, durante un control de identidad, se niegue a ser identificada o no pueda declarar o demostrar su identidad» incluso sin una autorización previa (considerando 33).

## **Los reversos (i)limitados del Reglamento de IA: sesgos y riesgos asociados a la discriminación algorítmica en contextos migratorios**

Los problemas de discriminación asociados al diseño del algoritmo y de la muestra de entrenamiento pueden tener su causa en cómo se definen la «variable objetivo» y las «etiquetas de clase», así como en cómo se realiza el etiquetado de los datos de entrenamiento, la recogida de los datos de entrenamiento, la selección de características y los *proxies* (FRA, 2022: 50); si bien la doctrina especializada

(Leese, 2014; Xenidis, 2020) y los principales informes no descartan una alta probabilidad de que se produzca un trato discriminatorio como resultado de la aplicación de tales sistemas (EC 2020a, 2020b y 2021b). Esta irrefrenable capacidad de elaboración, evaluación y predicción de perfiles e indicadores de riesgo por aproximación, adquieren una especial consideración cuando se pueden crear

**El despliegue tecnológico de sistemas basados en IA en el ciclo migratorio contribuye con mayor celeridad al etiquetado social y algorítmico de las personas migrantes bajo el pretexto de la seguridad, el miedo y la desconfianza.**

nuevos sesgos y reforzar estereotipos o prejuicios preexistentes hacia grupos históricamente marginados como las personas que cruzan fronteras regular o irregularmente y, en general, las personas migrantes. Por consiguiente, en este ámbito urge entender cuáles son las (im)posibilidades de revertir el trato discrimina-

torio y los procesos de estereotipización ya existentes, dados los nuevos retos que plantea la igualdad y la resiliencia del derecho antidiscriminatorio (Bartoletti y Xenidis, 2023; Morondo Taramundi, 2022).

#### a) Sesgos versus categorizaciones estereotipadas en los sistemas de IA aplicados en el contexto migratorio

La transformación *inteligente* de las fronteras exteriores de la UE y el despliegue tecnológico de sistemas basados en IA en el ciclo migratorio (Molnar, 2023: 309) contribuyen con mayor celeridad al etiquetado social y algorítmico<sup>14</sup> de las personas migrantes bajo el pretexto de la seguridad, el miedo y la desconfianza. Basta observar que tanto si un algoritmo contiene sesgos relacionados con una característica protegida, como si se omite la información de características protegidas como género, edad y origen étnico, no se puede garantizar totalmente que un sistema de IA no discrimine ni que haya varios ejes de discriminación/vulnerabilidad interseccional (Vantin, 2021). Al contrario, nada impide que pueda hacerlo de forma indirecta o por aproximación (*proxies*), incluso si no hay datos sensibles para elaborar «indicadores de riesgo». Por ejemplo, el sexo puede indicar por *aproximación* la orientación sexual; la ciudad de residencia puede revelar

---

14. Recomendación CM/Rec(2020)1 del Comité de Ministros a los estados miembros sobre las repercusiones de los sistemas algorítmicos en los derechos humanos (adoptada por el Comité de Ministros el 8 de abril de 2020 en la 1373ª reunión de los Delegados de los Ministros).

la etnia y el color; el nivel de estudios y la ocupación indicarán la propiedad y la afiliación sindical; la nacionalidad podría ser un indicador indirecto de origen étnico o religión, etc. (Vavoula, 2021; Brouwer, 2023).

Precisamente por ello, la inocuidad de las posibles categorías de sesgos no es una cuestión menor cuando la recogida y el tratamiento de *big data* perpetúa discriminaciones incluso de forma meramente informativa hacia determinados individuos o grupos en favor de otros (Bartoletti y Xenidis, 2023: 15-19). Además, estas categorías generales, si son aplicadas al ciclo migratorio (Beduschi, 2022; Suresh y Guttag, 2021), pueden adquirir otras significaciones. Por ejemplo, de un lado, la perpetuación de sesgos históricos, dada la discordancia entre el mundo real y los valores u objetivos que se codifican y propagan en el sistema de IA; el sesgo de representatividad, al definir y muestrear la población migrante sin una representación adecuada ni una generalización correcta para el entrenamiento y la evaluación del sistema de IA. De otro lado, el sesgo de medición, al escoger y medir las características y etiquetas que se utilizarán como indicadores indirectos de las magnitudes deseadas, pueden omitir algunos factores importantes, o introducir un *noise data* relacionado con los grupos o los datos de entrada que conduzca a resultados diferentes. Y, de igual modo, no plantean problemas menores aquellos sesgos que se producen durante fases de construcción, iteración o evaluación del sistema de IA, como son el sesgo de agregación cuando se combinan incorrectamente poblaciones heterogéneas que deberían tratarse por separado (migración económica «regular o irregular» y «solicitante de protección internacional»), o bien el sesgo de evaluación cuando la referencia externa no representa en igual medida a los distintos sectores de la población migrante utilizados para el entrenamiento y la evaluación, o se abren al uso de parámetros de comportamiento no adecuados (bueno/malo o falso/verdadero, sospechoso/criminal). Todo ello, sin perjuicio de la inevitable producción de un sesgo de despliegue que se manifiesta cuando un sistema de IA se utiliza o interpreta de maneras inadecuadas gracias a la interoperabilidad de los sistemas del Anexo x con fines policiales o preventivos.

En cualquier caso, Vantin (2021: 102) recuerda que el daño de los sesgos repercute sobre el grupo vulnerable, pero se extiende a los componentes de tal grupo, por lo que la desventaja, daño o prejuicio producido tendrá el efecto (por difracción) de consolidar o incrementar las discriminaciones estructurales o bien generar efectos desproporcionados sobre un grupo, más si cabe cuando se utilizan *sets* de datos de grandes dimensiones. De ahí que resulte extremadamente difícil, si no imposible, excluir la posibilidad de cualquier sesgo ni que se pueda garantizar en las mencionadas fases no causar efectos discriminatorios o nuevas formas de arbitrariedad, si las posibles limitaciones son delegadas a las autoridades responsables de diseñar o implementar un algoritmo y solo estas controlan el acceso a sus componentes técnicos esenciales.

Un primer reflejo concreto ya estudiado de la entidad práctica de estos efectos discriminatorios se produce, según Musco (2023: 45), cuando en el sistema de información ETIAS, la única limitación al desarrollador de algoritmos es permitir la elaboración de perfiles mediante la comparación de los indicadores de riesgo específico con los datos del expediente de solicitud (art. 33.4). Esta posibilidad deja margen para varias opciones de diseño no especificadas por el reglamento que, en la práctica, afectarán a la aplicación del algoritmo en el proceso de toma de decisiones, dada la prerrogativa soberana de los estados miembros de decidir quién entra en su territorio. Otro ejemplo de esta peligrosa tendencia tecnocéntrica, o tecnosolucionista, es la discrecionalidad técnica de la Agencia eu-LISA al considerar como «ruido» (*noise data*) las correlaciones y sesgos que arrojan tales sistemas de IA si tienen acceso o manejan información a gran escala, por ejemplo,

**Estas herramientas de IA pueden inferir, orientar y determinar automáticamente la designación o serialización de estereotipos para diferenciar entre los múltiples estatutos migratorios, añadiendo el factor de credibilidad/veracidad para etiquetarlos como buenos/malos migrantes, verdaderos/falsos solicitantes de asilo, vulnerables (in)vulnerables, etc.**

en situaciones de entrada irregular o identificación de vulnerabilidades en frontera (Nedelcu y Soysüren, 2022) o incluso, añadiría, con anterioridad a la llegada por medio del uso de «inocuos» *chatbots*.

Ahora bien, en otro orden de cosas, estos sesgos y el impacto discriminatorio que plantean los *proxie*, pese al marco europeo de IA, trascienden y pueden exacerbar los procesos de estereotipización que ya permean el régimen jurídico y político de la inmigración y el asilo.

Los estereotipos pueden responder a circunstancias o aspectos fácticos de las personas (y ser descriptivos), o a modelos de comportamiento (y ser en este sentido prescriptivos); en ambos casos, se establecen generalizaciones que no se detienen en recoger las peculiaridades individuales de los sujetos, sino que van más allá (Cook y Cusack, 2002). La categorización estereotipada de la movilidad humana (en posición regular, irregular, ya sea auténtica o falsa, etc.) puede llegar a ser criminalizada o etiquetada como una amenaza (Romano, 2023: 242) si tales herramientas de IA pueden inferir, orientar y determinar automáticamente la designación o serialización de estereotipos para diferenciar entre los múltiples estatutos migratorios, añadiendo el factor de credibilidad/veracidad para etiquetarlos como buenos/malos migrantes, verdaderos/falsos solicitantes de asilo, vulnerables (in)vulnerables, etc.

Así, la confianza en la fiabilidad algorítmica que se nutre del almacenamiento centralizado, entrenamiento y procesamiento automatizado de datos personales y biométricos pueden llegar a justificar *estadísticamente* la falsa conexión entre delincuencia e inmigración irregular para legitimar modelos punitivos, así como promover mayores medidas de lucha y prevención contra el terrorismo o el trá-

fico ilícito de personas. Esto es, el uso amplificado de la IA intensifica de facto la vigilancia de comportamientos *sospechosos* o *poco creíbles* de personas migrantes en los márgenes por medio de muestras sesgadas con implicaciones graves, no solo para las personas indocumentadas, sino también para aquellas documentadas que, ante la probabilidad de errores, además tienen difícil el acceso a la información sobre cómo se utilizan los datos y cómo pueden rectificarlos u obtener recursos efectivos en caso de abusos.

Al respecto, finalmente, cabe no olvidar, tal y como advierte Forti (2024: 3-4), que buena parte de los proyectos de aprendizaje automático o análisis predictivos basados en IA (véase tabla 1) pueden caer dentro de los denominados bucles de retroalimentación (*feedback loop*), que se producen cuando los resultados algorítmicos interpretan erróneamente la realidad e influyen en la aplicación de datos de entrenamiento para actualizar y desarrollar algoritmos. Los resultados engañosos y los estereotipos, además, pueden convertirse en entradas erróneas si se ajustan para otras rutinas de entrenamiento y solo buscan coincidencias exactas sin tener en cuenta las similitudes y correlaciones entre la información elaborada. Así pues, el resultado puede ser discriminatorio si la decisión adoptada basada en el sistema algorítmico conduce directa o indirectamente con efectos desproporcionados a un trato menos favorable incluso si está justificada objetiva y razonablemente para el fin que se emplea (Brouwer, 2020).

Por lo tanto, para determinar el alcance real de la discriminación algorítmica, un algoritmo debe evaluarse siempre dentro del contexto particular y la finalidad técnica de su uso (FRA, 2022: 24), admitiendo que, en contextos migratorios, su propósito persigue abiertamente predefinir las variables objetivo y las etiquetas de clase conforme a un marco normativo y político que no incluye *tout court* la condición migrante como patrón histórico de discriminación. De ahí que la única forma de garantizar estándares de protección vía Reglamento de IA no solo es prohibir aquella de origen étnico o racial sino «reducir la precisión general de todas las determinaciones y ampliar la diversidad de información» (Zuiderveen Borgesius, 2018: 55-60).

## b) Riesgos asociados a los sistemas de IA en el contexto migratorio: dinámicas asimétricas de poder y de injusticia social

Si bien es cierto que desde la UE se ha realizado de forma bienintencionada un mapeo de los riesgos de la IA sobre los derechos de las personas migrantes (EDRI, 2022; FRA, 2020), con ello no se ha frenado la capacidad de crear nuevas vulnerabilidades intrínsecas a este espacio arquetípico de dinámicas asimétricas de poder y de injusticia social. Las personas migrantes vulnerables son un colectivo que soporta la carga de los sistemas estatales de vigilancia banóptica y panóptica (Lee-

se *et al.*, 2022: 6-7) ahora reconvertidos en «arquitecturas carcelarias» (Molnar, 2019). No solo para conducirlos hacia formas de movilidad (negocio o turismo) que se alientan o facilitan por su conformidad al sistema jurídico-político de entrada/salida, sino también para inhibir o desacelerar aquella espiral de amenazas asociadas a la «migración irregular», al «falso asilo» o a la criminalidad migrante (Leese, 2014). Por tanto, según Okzul (2023: 15), las implicaciones de la geopolítica de la movilidad humana asociadas a estas herramientas de IA tienen una doble vertiente dispar: de un lado, ayudar a las autoridades de inmigración e incluso a organizaciones humanitarias a planificar y asignar mejor sus recursos de antemano; y, de otro lado, aumentar la vigilancia de las fronteras y respaldar políticas de contención, expulsión y/o exclusión que restrinjan la migración en lugar de facilitarla, por medio de decisiones discrecionales (Molnar, 2019: 56) en defensa

**La puesta en marcha de proyectos experimentales ya muestra evidencias sobre la exposición de las poblaciones sometidas arbitrariamente a la violencia estatal por su condición de personas migrantes racializadas, mediante una creciente vigilancia biotecnológica/biopolítica de control securitario sobre la movilidad humana no querida.**

de la soberanía nacional amenazada o para mantener los privilegios entorno al conocimiento y las competencias de preconfiguración algorítmica de datos (ibídem, 2023: 48 y 49).

Así pues, sobre el alcance del riesgo en función de los diferentes actores implicados y sus objetivos, es necesario centrarse en el desequilibrio de posiciones de poder entre países con capacidad de IA o sin ella, esto es, los

llamados «AI-capable» o «AI-dependent» (Beduschi, 2021: 4), y en la precariedad social y política (*precarity*) (Butler, 2009: 2) como «forma de agresión que no es inducida por el Estado, pero contra la que este no proporciona una protección adecuada». La puesta en marcha de proyectos experimentales como los mencionados de carácter preventivo, predictivo, procedimental o informativo, ya muestra evidencias sobre la exposición de las poblaciones sometidas arbitrariamente a la violencia estatal por su condición de personas migrantes racializadas, mediante una creciente vigilancia biotecnológica/biopolítica de control securitario sobre la movilidad humana no querida. Esta desigualdad y asimetría de poderes genera conflictos de intereses si el uso de la IA recae principalmente en el sector privado (IDEMIA; INDRA; SenseTime) que diseña y desarrolla tecnologías de IA que utilizarán los estados, las agencias y las organizaciones internacionales en todo el ciclo migratorio. Por tanto, estas interacciones entre los sectores público y privado suscitan inquietud no solo por la quiebra de la protección de datos y el principio de proporcionalidad (Jasmontaite-Zaniewicz y Zomignani, 2021), sino porque pueden tener mayor trasfondo en la producción unilateral de la «verdad o la credibilidad absoluta» sobre las personas migrantes. Por ejemplo, no podemos olvidar

cómo, en los procesos automatizados de toma de decisiones relacionados con la migración y el asilo, a menudo, los derechos de autor y de propiedad intelectual de los creadores de sistemas de IA pueden obstaculizar la difusión de información sobre la creación de los algoritmos o el modelo de entrenamiento, lo que genera mayor inseguridad jurídica y falta de confianza sobre cómo y para qué pueden utilizarse o no los datos recopilados y sus correlaciones derivadas.

Por último, otro motivo añadido de creciente preocupación por este monopolio y desequilibrio de poder a nivel macro es advertido desde los debates sobre el «tecnocolonialismo» y el «colonialismo de datos» (Mejias y Couldry, 2019). No en vano, la apropiación con IA de datos y de muestras sesgadas abre la brecha del paradigma Norte y Sur Global, no solo bajo una fachada de objetividad y modernización amigable, sino porque puede facilitar nuevas formas de sobrexplotación, racismo y dominación<sup>15</sup>, especialmente hacia mujeres, minorías étnicas, personas con discapacidad y colectivos LGTBI. Estos colectivos integran a las personas más conocibles, experimentales, gobernables, rastreables e inteligibles (Molnar, 2019) y, según Romano (2023: 243), ello entraña una reificación o degradación del ser humano a un mero objeto. Esto es, comporta una pérdida de su autonomía y de su dignidad humana al ser tratados como datos «medio-fin» sin valor que se pueden filtrar, correlacionar, desechar o (re)usar sine die.

## Algunas breve consideraciones

Si bien la UE se presenta como un escenario receptivo para proteger los derechos de las personas migrantes, al considerar su vulnerabilidad como una prioridad en la agenda europea, su trascendencia puede ser relativa, básicamente si se consideran los sesgos y riesgos asociados al creciente despliegue de sistemas tecnológicos basados en IA en contextos de control y gestión migratoria tras la plena vigencia y aplicación del Reglamento de IA. De hecho, no solo hay numerosas evidencias que dudan de su alcance prometedor, sino que en tales contextos, las mayores incertidumbres que plantean las prohibiciones y usos admisibles de IA en comparación con las que resuelve, niegan toda posible interpretación *pro personae* del Reglamento. No en vano, aunque la introducción progresiva de IA podría perseguir no solo fines

---

15. Comité Asesor del Consejo de Derechos Humanos de Naciones Unidas. «Promover la justicia y la igualdad raciales desarraigando el racismo sistémico» (8 de agosto de 2023), UN Doc. A/HRC/54/70, párr. 7.

punitivos sino también fines constructivos (operaciones humanitarias de salvamento, el análisis rápido y eficaz de las solicitudes de asilo y de la información, o logística en procesos de acogida-integración), la opacidad inherente al funcionamiento algorítmico, la imposibilidad de reducción del riesgo, el inevitable impacto discriminatorio de los sesgos y de los *proxies* (que no se mencionan) y la falta de garantías sólidas de los derechos en juego, no colman los vacíos de protección exigibles.

Además, tampoco los esfuerzos que se desprenden del Reglamento de IA parecen orientarse a evitar la previsible difracción algorítmica de sesgos y estereotipos de la aplicación de la IA, dado que ni tan siquiera se renuncia como «mal menor» a minimizar las repercusiones negativas que implica excluir los sistemas del Anexo x, especialmente para los grupos migrantes más vulnerables. Ello es difícilmente salvable si los regímenes de migración y de asilo tienen como base categorías migrantes estereotipadas y se asientan en el poder soberano de atribuir/cambiar espacios de libertad versus seguridad controlada, pero también si tales sistemas tienen el poder de responsabilizar sin atribuir responsabilidad o agencia a las personas migrantes simplemente para mantener inalterada la legitimidad o el statu quo del propio sistema. Así pues, por medio de justificaciones objetivas, excepciones o prórrogas (i)limitadas, la condición migrante sigue siendo relegada a los márgenes de los patrones históricos de discriminación de los sistemas jurídico-políticos y también ahora de los sistemas tecnológicos de IA, lo que crea escenarios migratorios y confines fronterizos cada vez más distópicos.

## Referencias bibliográficas

- Akhmetova, Roxana. «Efficient Discrimination: On How Governments Use Artificial Intelligence in the Immigration Sphere to Create and Fortify ‘Invisible Border Walls’». *COMPAS*, Working Paper 149, (2020), p. 1-23 (en línea) [https://www.compas.ox.ac.uk/wp-content/uploads/WP-2020-149-Akhmetova\\_Efficient\\_Discrimination.pdf](https://www.compas.ox.ac.uk/wp-content/uploads/WP-2020-149-Akhmetova_Efficient_Discrimination.pdf)
- Avello Martínez, María. «EU Borders and Potential Conflicts between New Technologies and Human Rights». *Peace & Security – Paix et Sécurité Internationales*, n.º 11, (2023), p. 1-33.
- Bartoletti, Ivana y Xenidis; Raphaële. *Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality and the risks they may cause in relation to discrimination*. Estrasburgo: Council of Europe, 2023.
- Beduschi, Ana. «International migration management in the age of artificial intelligence». *Migration Studies*, vol. 9, n.º 3 (2021), p. 576-596.

- Brouwer, Evelien. «*Challenging Bias and Discrimination in Automated Border Decisions: Ligue des droits humains and the Right to Effective Judicial Protection*». *VerfBlog*, (11 de mayo de 2023) (en línea) <https://verfassungsblog.de/pnr-border/>
- Beduschi, Ana. *Artificial intelligence, migration and mobility: implications for policy and practice*. Ginebra: World Migration Report, 2022.
- Broeders, Dennis. «The New Digital Borders of Europe: EU Databases and the Surveillance of Irregular Migrants». *International Sociology*, vol. 22, n.º 1 (2007), p. 71-92.
- Burgess, J. Peter y Kloza, Dariusz. *Border control and new technologies. Addressing integrated Impact Assessment*. Bruselas: Academic and Scientific Publishers, 2021.
- Butler, Judith. «Performativity, precarity and sexual politics». *AIBR. Revista de Antropología Iberoamericana*, vol. 4, n.º 3 (2009), p. 1-13.
- Chander, Anupam. «The racist algorithm?». *Michigan Law Review*, vol. 115, n.º 6 (2017), p. 1.023-1.045.
- Cook, Rebecca J. y Cusack, Simone. *Gender Stereotyping: Transnational Legal Perspectives*. Filadelfia: University of Pennsylvania Press, 2010.
- Cruz Ángeles, Jonatán. «Procesamiento informático de datos y protección de derechos fundamentales en las fronteras exteriores de la Unión Europea». *Freedom, Security and Justice: European legal studies*, n.º 1 (2020), p. 94-122.
- Cruz Ángeles, Jonatán. «La nueva Ley de Inteligencia Artificial una aliada necesaria para gestionar controles fronterizos, migratorios y de asilo en la Unión Europea», en: Vestri, Gabriele (dir.) *La disrupción tecnológica en la Administración Pública: retos y desafíos de la inteligencia artificial*. Cizur Menor: Aranzadi, 2022, p. 97-121.
- Garrido Carrillo, Francisco Javier. «La inteligencia artificial en el control de los flujos migratorios en la Unión Europea. La necesidad de un marco normativo garantista de los derechos fundamentales». *Revista General de Derecho europeo*, n.º 60 (2023), p. 1-39.
- Derave, Charly; Genicot, Nathan y Hetmanska, Nina. «The Risks of Trustworthy Artificial Intelligence: The Case of the European Travel Information and Authorisation System». *European Journal of Risk Regulation*, vol. 13, n.º 3 (2022), p. 389-420.
- Dumbrava, Costica. *Artificial intelligence at EU borders. Overview of applications and key issues*. Bruselas: European Parliamentary Research Service, 2021.
- EC-European Commission. *Communication from the Commission to the European Parliament and the Council on Smart borders: options and the way forward*. Bruselas: EC, 2011.
- EC-European Commission. *Communication from the Commission to the European Parliament and the Council on Stronger and Smarter Information Systems for Borders and Security*. Bruselas: EC, 2016.

- EC-European Commission. *The Use of Artificial Intelligence in Border Control, Migration and Security*. Bruselas: EC, 2020a.
- EC-European Commission. *White paper on Artificial Intelligence - A European approach to excellence and trust*. Bruselas: EC, 2020b.
- EC-European Commission. *Artificial Intelligence in EU Borders*. Bruselas: EC, 2021a.
- EC-European Commission. *Feasibility study on an early warning and forecasting tool for migration based on artificial intelligence technology*. Bruselas: EC, 2021b.
- EC-European Commission. *Opportunities and Challenges for the Use of Artificial Intelligence in Border Control, Migration and Security*. Bruselas: EC, 2021c.
- EDRI-European Digital Rights. *Uses of AI in immigration and border control. A fundamental rights approval to the Artificial Intelligence Act*. Bruselas: EDRI, 2022.
- EMN-European Migration Network. *The use of digitalization and artificial intelligence in migration management*. Bruselas: EMN, 2022.
- FRA-European Union Agency for Fundamental Rights. *Facial recognition technology: fundamental rights considerations in the context of law enforcement*. Viena: FRA publications, 2020.
- FRA-European Union Agency for Fundamental Rights. *Bias in Algorithms – Artificial Intelligence and Discrimination*. Viena: FRA publications, 2022.
- Forti, Mirko. «Addressing Algorithmic Errors in Data-Driven Border Control Procedures». *German Law Journal*, vol. 25, n.º 1 (2024), p. 1-11.
- Guild Elspeth y Vavoula, Niovi. «Travel Authorization in the EU: Automated Processing and Profiling». *OpenDemocracy*, (12 de octubre de 2020) (en línea) <https://www.opendemocracy.net/en/can-europe-makeit/travel-authorization-eu-automated-processing-and-profiling/>
- Jasmontaite-Zaniewicz, Lina y Zomignani Barboza, Júlia. «Disproportionate surveillance: technology assisted and automated decisions in asylum applications in the EU?». *International Journal of Refugee Law*, n.º 33 (2021), p. 89-110.
- Kilpatrick, Jane y Jones, Chris. *A clear and present danger. Missing safeguards on immigration and asylum EU's AI act*. Londres: Statewatch, 2022.
- Leese, Mathias. «The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union». *Security Dialogue*, vol. 45, n.º 5 (2014), p. 494-511.
- Leese, Mathias; Noori, Simon y Scheel, Stephan. «Data Matters: The Politics and Practices of Digital Border and Migration Management». *Geopolitics*, 27, n.º 1 (2022), p. 5-25.
- Loddo, Olimpia Giuliana y Addis, Andrea. «Tre funzioni dell'IA nella gestione dei flussi migratori». *Adimblog, Analisi & Opinioni*, (enero de 2023) (en línea) <https://www.adimblog.com/2023/01/31/tre-funzioni-dellia-nella-gestione-dei-flussi-migratori/>

- Marin, Luisa. «Is Europe Turning into a ‘Technological Fortress’? Innovation and Technology for the Management of EU’s External Borders», en: Heldeweg, Michiel A. y Kica, Evisa (eds.) *Regulating Technological Innovation: Legal and Economic Regulation of Technological Innovation*. Londres: Palgrave Macmillan, 2011, p. 131-151.
- Mejías, Ulises y Couldry, Nick. «Colonialismo de datos: repensando la relación de los datos masivos con el sujeto contemporáneo». *Virtualis*, vol. 10, n.º 18 (2019), p. 78-97.
- Molnar, Petra. «Technology on the Margins: AI and Global Migration Management from a Human Rights Perspective». *Cambridge International Law Journal*, 8, n.º 2 (2019), p. 305-330.
- Molnar, Petra. «Digital border technologies, techno-racism and logics of exclusion». *International Migration*, vol. 61, n.º 5 (2023), p. 307-312.
- Molnar, Petra. *AI surveillance and border tech experiments*. Nueva York: The New Press, 2024.
- Moreno Rebato, Mar. *Inteligencia artificial (umbrales éticos, derecho y administraciones públicas)*. Cizur Menor: Aranzadi, 2021.
- Morondo Taramundi, Dolores. «Discrimination by Machine-Based Decisions: Inputs and Limits of Anti-discrimination Law», en: Custers, Bart, Fosch-Villaronga; Eduard (eds.) *Law and Artificial Intelligence*. La Haya: T.M.C. Asser Press/Springer, 2022, p. 73-85.
- Musco Eklund, Amanda. «Rule of Law Challenges of ‘Algorithmic Discretion’ & Automation in EU Border Control. A Case Study of ETIAS Through the Lens of Legality». *European Journal of migration and Law*, vol. 25, n.º 3 (2023), p. 249-274.
- Nalbandian, Lucia. «An eye for an ‘I’: a critical assessment of artificial intelligence tools in migration and asylum management». *Comparative Migration Studies*, vol. 10, n.º 32 (2022), p. 1-23.
- Napolitano, Antonella. *Artificial intelligence: the new frontier of the EU’s border externalisation strategy*. Copenhagen: EuroMed rights, 2023.
- Nedelcu, Mihaela y Soysüren, Ibrahim. «Precarious migrants, migration regimes and digital technologies: the empowerment-control nexus». *Journal of Ethnic and Migration Studies*, vol. 48, n.º 8 (2022), p. 1.821-1.837.
- Oliveira Martins, Bruno y Gabrielsen Jumbert, Maria. «EU Border technologies and the co-production of security ‘problems’ and ‘solutions’». *Journal of Ethnic and Migration Studies*, vol. 48, n.º 6 (2022), p. 1.430-1.447.
- Ozkul, Derya. *Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe*. Oxford: Refugee Studies Centre, University of Oxford, 2023.

- Romano, Andrea. «Derechos fundamentales e inteligencia artificial emocional en iBorderCtrl: retos de la automatización en el ámbito migratorio». *Revista Catalana de Dret Públic*, n.º 66 (2023), p. 237-252.
- Ruiz Tarrías, Susana. «La búsqueda del modelo regulatorio de la IA en la Unión Europea». *Anales de la Cátedra Francisco Suárez*, n.º 57 (2023), p. 91-119.
- Suresh, Harini y Gutttag, John. «A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle», en: *Proceedings of EAAMO '21: Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21)*, (octubre de 2021) (en línea) <https://dl.acm.org/doi/fullHtml/10.1145/3465416.3483305>
- UN Special Rapporteur on Contemporary forms of racism, racial discrimination, xenophobia and related intolerance. «Racial and xenophobic discrimination, emerging digital technologies, and border and immigration enforcement». UN, A/HRC/44/57, (18 de junio de 2020) (en línea) <https://digitallibrary.un.org/record/3879751?v=pdf>
- Vantin, Serena. *Il diritto antidiscriminatorio nell'era digitale. Potenzialità e rischi per le persone, La Pubblica Amministrazione, le imprese*. Milán: Wolters Kluwer, 2021.
- Vavoula, Niovi. «Artificial Intelligence (AI) at Schengen Borders: Automated Processing, Algorithmic Profiling and Facial Recognition in the Era of Techno-Solutionism». *European Journal of Migration and Law*, n.º 23 (2021), p. 457-484.
- Veale, Michael y Zuiderveen Borgesius, Frederik. «Demystifying the Draft EU Artificial Intelligence Act». *CRi*, n.º 4 (2021), p.1-16.
- Xenidis, Raphaël. «Tuning EU equality law to algorithmic discrimination: Three pathways to resilience». *Maastricht Journal of European and Comparative Law*, vol. 27, n.º 6 (2020), p. 736-758.
- Zuiderveen Borgesius, Frederik. *Discrimination, artificial intelligence, and algorithmic decision-making*. Estrasburgo: Council of Europe, Directorate General of Democracy, 2018.

# La llegada de la IA a las ciudades: urbanismo de plataformas y nuevos regímenes de gubernamentalidad

## The arrival of AI to the cities: platform urbanism and new regimes of governmentality

### José Luis Blasco Ejarque

Investigador predoctoral, Universitat Autònoma de Barcelona (UAB).  
joseluis.blasco@autonoma.cat. ORCID: <https://orcid.org/0000-0001-7675-3349>

### Francisco Javier Tirado

Profesor titular, Departamento de Psicología Social, Universitat Autònoma de Barcelona (UAB).  
franciscojavier.tirado@uab.cat. ORCID: <https://orcid.org/0000-0001-7093-056X>

**Cómo citar este artículo:** Blasco Ejarque, José Luis y Tirado, Francisco Javier. «La llegada de la IA a las ciudades: urbanismo de plataformas y nuevos regímenes de gubernamentalidad». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 195-217. DOI: doi.org/10.24241/rcai.2024.138.3.195

**Resumen:** Este artículo analiza el denominado «urbanismo de plataformas» como un dispositivo *posbiopolítico* de dominación y subjetivación propio de las «sociedades de rendimiento» del siglo XXI. En tal sentido, se examina cómo las plataformas algorítmicas exponen a la ciudadanía a nuevas relaciones de control y poder a través de dispositivos tecnosociales como la inteligencia artificial (IA). Los autores concluyen con la idea de que el despliegue del urbanismo de plataformas responde a un modelo de «gubernamentalidad algorítmica» que propicia el advenimiento de nuevas prácticas de disidencia basadas en la denominada «innovación abierta».

**Palabras clave:** capitalismo de plataforma, urbanismo de plataformas, inteligencia artificial (IA), biopolítica, psicopolítica, ciudades inteligentes o *smart cities*

**Abstract:** This paper analyses “platform urbanism” as a post-biopolitical *dispositive of domination and subjectivation* characteristic of the “performance societies” of the 21st century. It examines how algorithmic platforms expose citizens to new control and power relations through techno-social dispositives such as artificial intelligence (AI). The authors close with the idea that the rollout of platform urbanism is a model of “algorithmic governmentality” that fosters the advent of new practices of dissent based on “open innovation”.

**Key words:** platform capitalism, platform urbanism, artificial intelligence (AI), biopolitics, psychopolitics, smart cities

*El presente artículo se desprende de la investigación «Escenarios de futuro de las políticas de desarrollo local en la época postpandemia. De la biopolítica a la psicopolítica», que forma parte de un proyecto general denominado «Las relaciones de poder e innovación abierta». El trabajo se ha realizado en el marco de investigación del grupo Barcelona Science and Technology Studies (STSb), programa de doctorado «Persona y Sociedad en el Mundo Contemporáneo», Departamento de Psicología Social, Universitat Autònoma de Barcelona.*

Desde finales de la pasada centuria, la digitalización, el uso de sensores, robots y algoritmos se ha generalizado en todos los ámbitos de nuestra vida cotidiana. Del simple despertador que nos alerta por las mañanas hasta el automóvil que utilizamos para desplazarnos, toda nuestra experiencia diaria está sembrada de elementos tecnológicos vinculados de algún modo con la IA. Las herramientas de gestión de datos masivos y predicción, de toma de decisiones automatizada, de reconocimiento visual y auditivo, de registro de datos corporales, etc., se han convertido en elementos de un paisaje que cada vez es más habitual y natural. Ya no nos sorprende la contratación automática, los almacenes inteligentes, las estructuras algorítmicas que controlan los procesos de selección de personal o la presencia de robots en la producción de todo tipo de manufacturas.

La mencionada revolución ha impactado de un modo especial en el mundo de las organizaciones (grandes compañías y pequeñas empresas), del trabajo y de la educación. Sus procesos de gestión y producción, productos, maneras de operar y relacionarse cambian a una velocidad de vértigo. El uso cotidiano de *smartphones* y ordenadores de todo tipo en el espacio laboral, el análisis de fuentes de datos y de redes sociales en la educación, la implementación de calendarios compartidos o instrumentos de trabajo colaborativo, así como la aparición de sistemas de IA para controlar y monitorizar toda esa actividad han inaugurado una revolución en nuestra cotidianidad que todavía no está muy claro hacia qué direcciones apuntará. Eso sí, pone de manifiesto que el cambio está aquí y es insoslayable.

Ese cambio también afecta a nuestras urbes, ya que las ha comenzado a transformar de tal manera que algunos conceptos que hemos utilizado en su análisis y comprensión se han quedado obsoletos. Las ciudades del siglo XXI dependen cada vez más de modelos predictivos fundamentados en datos para la toma de decisiones automatizada, los cuales se vinculan directamente con el desarrollo de algoritmos y de la IA. Su implementación masiva en entornos urbanos anuncia lo que se ha denominado «urbanismo de plataformas» (Van der Graaf y Ballon, 2019) y nos prepara para el advenimiento de un nuevo régimen de gubernamentalidad (Castro Gómez, 2015) que presenta, entre otras cosas, nuevas formas de agencia y de relaciones de control y poder. Así, la «datificación» de la vida urbana a través del desarrollo de ciudades gemelas digitales<sup>1</sup> (Deng *et al.*, 2021), la simulación de sociedades artificiales o la apro-

---

1. Gemelo digital hace referencia a la reproducción en un entorno virtual de una organización, institución o ciudad.

piación del entorno urbano por parte de corporaciones empresariales digitales, son algunos ejemplos que avalan lo mencionado y, además, introducen nociones como la de *desigualdad algorítmica*, *reapropiación ciudadana* y *disidencia* como la cara oculta del mencionado «urbanismo de plataforma» (Caprotti y Liu, 2022; Morozov, 2016).

Esta imparable *plataformización* de las urbes no ha dejado indiferente a las distintas disciplinas sociales y tecnológicas. En ese sentido, los análisis que se proponen se pueden agrupar, de un modo muy esquemático, en dos grandes constelaciones: por un lado, encontramos propuestas posibilistas que visualizan la ciudad como una gigantesca plataforma que optimiza y potencia recursos para el ciudadano (Townsend, 2013; Herzberg, 2017); por otro, observamos estudios críticos que analizan las prácticas de agencia que se vinculan con los artefactos sociotécnicos en los que se fundamentan las plataformas, por ejemplo, algoritmos o bases de datos (Rossi, 2022). Dichas plataformas encierran un nuevo tipo de relaciones de control y poder que, en ocasiones, recibe la denominación de gubernamentalidad algorítmica (Rouvroy y Berns, 2013) o dispositivos *posbiopolíticos* (Han, 2014).

Este artículo se enmarca en el segundo conjunto de estudios y entiende el «urbanismo de plataformas» como un dispositivo que genera una intersección específica entre sociedad, tecnología, capital y territorio. Por lo tanto, el análisis y la problematización del mencionado urbanismo supone desvelar qué transformaciones se generan en esos elementos y qué nuevos tipos de relaciones políticas y de poder se conforman. En ese sentido, este trabajo sostiene la hipótesis de que esa peculiar intersección tiene las características de lo que algunos autores (Han, 2014; Sloterdijk, 2014) han denominado *psicopolítica*. Para argumentar esta afirmación, se examinan tres realidades que operan al unísono: a) el urbanismo de plataformas supone la creación de relaciones de poder y control que, a su vez, configuran nuevos tipos de agencia para el ciudadano; b) las mencionadas relaciones de poder pasan desapercibidas porque no exhiben el formato de las ampliamente reconocidas relaciones biopolíticas y, en su lugar, conforman lo que denominamos dispositivos *posbiopolíticos* o *psicopolíticos*; y c) aparecen nuevas arquitecturas de disidencia (Sadín, 2023).

**Tres realidades operan al unísono: a) el urbanismo de plataformas supone la creación de relaciones de poder y control que, a su vez, configuran nuevos tipos de agencia para el ciudadano; b) las mencionadas relaciones de poder pasan desapercibidas porque no exhiben el formato de las ampliamente reconocidas relaciones biopolíticas y, en su lugar, conforman lo que denominamos dispositivos *posbiopolíticos* o *psicopolíticos*; y c) aparecen nuevas arquitecturas de disidencia.**

Para realizar este ejercicio, el artículo se estructura de la siguiente forma. En primer lugar, se exponen las actuales narrativas tecnosociales que imperan en nuestro imaginario social sobre la figura de la *urbe*, las cuales, básicamente, se resumen en las nociones de «urbanismo inteligente» y «capitalismo de plataforma». La sección se cierra argumentando que ambas narrativas anuncian, en términos socioculturales, la conformación de un nuevo tipo de sociedad que algunos autores denominan «sociedad del rendimiento» (Han, 2014). En segundo lugar, tras constatar que en tal sociedad se impone un nuevo régimen de gubernamentalidad que opera a partir del uso masivo e intensivo de algoritmos e IA, se sostiene que dicho régimen permite entender cómo las denominadas «ciudades inteligentes» (*smart cities*) se están convirtiendo progresivamente en «ciudades posinteligentes». En tercer lugar, se examina cómo en tales ciudades se constituye la disidencia –y la forma que esta puede adquirir– frente a la nueva gubernamentalidad. Por último, se concluye señalando que las plataformas se pueden analizar desde diversas perspectivas: como espacio físico y virtual, como una relación laboral, como un sistema de sujeción... Sin embargo, su caracterización como dispositivos, en el sentido que Michel Foucault dio al término, permite visualizar dimensiones que de otra manera quedan ocultas. Por ejemplo, se muestra qué tipos de regímenes y gubernamentalidad se establecen en las urbes del siglo XXI y qué posibilidades de resistencia-disidencia se tornan factibles.

## La urbe y sus narrativas tecnosociales

Hace ya algunas décadas, la obra de William J. Mitchell *City of bits: Space, place and the infobahn* (1996) se convirtió en un clásico de la literatura académica. En ella aparecía la evidencia de que el advenimiento de la era digital transfería al espacio virtual la casi totalidad de las actividades sociales, laborales, económicas, culturales y urbanísticas. Por consiguiente, el nuevo propósito del urbanismo debía ser el desarrollo de interfaces digitales que constituyesen metrópolis virtuales. Esta idea incubó el nacimiento de nociones como *smart cities* o *living labs*, ambas representando modelos de experimentación en entornos de la vida real y nacidas con la vocación de observar la denominada «interacción computadora-humano» (*human-computer interaction* [HCI]) (véase Kashef *et al.*, 2021).

A lo largo de la década de los años ochenta del siglo pasado, el desarrollo y madurez de tecnologías de conectividad alámbrica e inalámbrica (LAN, WAN, Wi-Fi) facilitaron la aparición de las llamadas *smart cities*, un concepto que fue difundido y universalizado por corporaciones como IBM que,

en su publicación *A vision of smarter cities. How cities can lead the way into a prosperous and sustainable future* (Dirks *et al.*, 2010), menciona en múltiples ocasiones. Estas corporaciones estaban decididas a desarrollar un conjunto de tecnologías ubicuas que facilitasen la tecnificación de las ciudades con el propósito de mejorar las infraestructuras públicas y optimizar la toma de decisiones de los gobiernos municipales, dotándoles de indicadores objetivos sobre la evolución de la ciudad en tiempo real. Por esta razón, la *smartización* de la ciudad siguió, de alguna manera, la lógica de los modelos productivos de las factorías y empresas de *hardware* y *software*: Cisco, Siemens, Microsoft, Intel, IBM o Hitachi, entre otras, que transformaron las urbes en factorías sociotécnicas a partir de una industria emergente que combinaba el desarrollo de tecnologías virtuales con tecnologías auxiliares 4.0<sup>2</sup> que permitían la denominada «datificación». Ello supuso también la datificación masiva de las ciudades y la aparición del llamado «diluvio digital» (Cortada, 2012) o la «revolución de los datos» (Kitchin, 2014).

El actual interés por los procesos algorítmicos se correlaciona directamente con la ingente producción de datos digitales generados a través del desarrollo de diferentes tecnologías *smart*. En tal producción juegan un papel muy relevante tanto las tecnologías basadas en sensores del entramado urbano (*smart city*), como las informacionales, comunicativas o metamediáticas generadas por los usuarios a través de sus propios *smartphones*. De esta forma, en el proceso de datificación, los algoritmos son los encargados de gestionar este inmenso flujo de datos; aunque también, y quizás de forma más relevante, de trasladar este conjunto de gramáticas asignificantes<sup>3</sup> (Deleuze y Guattari, 1972) a dispositivos con capacidad de agencia y propiciar la emergencia de dos sistemas sociotécnicos: el identificado como «urbanismo inteligente» y el denominado «capitalismo de plataforma». Estos serían los antecedentes contemporáneos del ubicuo e inexplorado urbanismo IA desarrollado en las denominadas «ciudades posinteligentes» (Cugurullo *et al.*, 2023).

- 
2. Las tecnologías 4.0 hacen referencia a dispositivos como el Internet de las Cosas Industrial (IIoT, por sus siglas en inglés), la conectividad en la nube o la IA y *machine learning*. Este enfoque unificado e integrado de la fabricación da como resultado productos, fábricas y activos conectados e inteligentes.
  3. Estos autores definen las gramáticas asignificantes por oposición a las semióticas significantes. Mientras que las segundas operan generando significados, a través del uso de códigos numéricos, alfabéticos, etc., las primeras hacen referencia al uso de señales y elementos gráficos como tablas, bases de datos, etc., que operan a partir del ejercicio de su mera circulación y conexión con otros códigos. Es decir, no tienen como principal objetivo la creación de un significado, sino simplemente el hecho de circular como meros significantes y provocar efectos con su mero movimiento.

## Urbanismo inteligente

El urbanismo inteligente se presenta con el objetivo de mejorar la eficiencia de los sistemas urbanos a través de tecnologías *smart* ubicadas en estructuras denominadas «ciudades inteligentes». Dichas tecnologías han sido desarrolladas por grandes corporaciones y desplegadas en las metrópolis gracias a la colaboración de los gobiernos locales. Bajo esta lógica, las autoridades municipales se han transformado en el principal valedor de las actividades de las ciudades inteligentes; así, la financiación de proyectos tecnológicos ha desplazado la tipología de inversión de los entes públicos, desarrollándose una fuerte relación entre los municipios y las consultoras tecnológicas privadas (Mazzucato y Collington, 2023). Por lo tanto, no resulta arriesgado afirmar que la ciudad inteligente se desarrolla

**El urbanismo inteligente se presenta con el objetivo de mejorar la eficiencia de los sistemas urbanos a través de tecnologías *smart* ubicadas en estructuras denominadas «ciudades inteligentes». Dichas tecnologías han sido desarrolladas por grandes corporaciones y desplegadas en las metrópolis gracias a la colaboración de los gobiernos locales.**

a partir de la financiación pública y se articula a través del prisma tecnoliberal de consultoras y corporaciones privadas. Un ejemplo muy claro son iniciativas como WiFi4EU (Comisión Europea, 2017), que pretendía tener un impacto concreto en los municipios de toda Europa al promover el acceso gratuito a la conectividad Wi-Fi para la ciudadanía en los espacios públicos. La tecnoinnovación en las ciudades inteligentes se presenta con el propósito de facilitar la gestión municipal, convertir la administración en una eficaz prestadora de servicios y renovar la relación entre los servidores públicos y la ciudadanía; prometiendo una evaluación de la gestión administrativa para mejorar el funcionamiento de las corporaciones locales, así como la orientación de la organización y servicios hacia el ciudadano.

La crisis de las hipotecas subprime (2007-2008) y el posterior *austericidio* redujeron los recursos de las economías locales y la disciplina fiscal aceleró la búsqueda de herramientas que consiguiesen menguar la capacidad de gasto público. De este modo, las tecnologías al servicio del capital ayudaron al crecimiento de una nueva industria *smart* que —financiada con fondos públicos y alimentada a través de ingentes cantidades de capital riesgo— aparentemente invertía en el sector tecnológico. Sin embargo, dicha industria realmente seguía invirtiendo en la urbanización de las ciudades, financiando nuevas capas de valorización digital sobre el territorio que permitían fomentar de nuevo una economía geoespacial, en este caso no fundamentada en el capitalismo financiero que llevó a las crisis hipotecarias, sino en un capitalismo de platafor-

ma que seguía operando espacialmente sobre las urbes. Dicho de otra forma: se desplegó una lógica que iba *from the brick to the byte* (del ladrillo al byte). Esta transformación vino acompañada de un cambio en los modelos de gobernanza. Y, así, se impuso en poco tiempo la participación ciudadana como modelo ideal de gobierno, un modelo que imprimía un giro participativo en las decisiones públicas a través de estrategias «de abajo hacia arriba» (*bottom-up*). Este escenario supuso la entrada directa en las denominadas «sociedades de rendimiento» (Han, 2013). En ellas, la ciudadanía aparece equipada con dispositivos tecnológicos que la convierten de facto en un agente de cogobierno, un activista o como el motor de la toma de ciertas decisiones. A través de *hackatones*, *crowdsourcing*, movilizaciones *online* o por medio de dispositivos inteligentes como los *smartphones* y las plataformas de gobierno abierto, la ciudadanía entró de lleno en la definición y gestión de sus ciudades y pasó a ser un agente más del denominado «capitalismo de plataforma».

## Capitalismo de plataforma

Para Srnicek (2017), la etiqueta «capitalismo de plataforma» señala una mutación del paradigma del capitalismo a partir del desarrollo de nuevas tecnologías adheridas a una lógica de extracción de datos. El concepto señala un nuevo proceso de acumulación del capital propio de las sociedades occidentales avanzadas en las que plataformas monopolísticas se convierten en el principal engranaje sociotécnico de las economías tecnoliberales del siglo XXI. La capacidad apropiativa de este nuevo capitalismo se entiende a través del desarrollo y ubicuidad de las tecnologías algorítmicas, omnipresentes en la mayoría de los ámbitos de la vida cotidiana a través de corporaciones supranacionales. El autor describe cinco tipos de plataformas digitales: a) las que centran su uso en la gestión de publicidad y monetizan datos personales a través de publicidad comportamental (Google y Facebook); b) las que ofrecen servicios de alojamiento en la nube destinadas al uso de terceros (Google Cloud Platform, Alibaba, IBM Cloud) y que son poseedoras de los mayores centros de procesamiento de datos (CPD), donde se alojan las plataformas que gestionan las urbes, por tanto los centros sinápticos del urbanismo *smart* y del novedoso urbanismo IA; c) las plataformas de origen industrial que digitalizan procesos productivos (sería el caso de las gemelas digitales que tienen corporaciones como Volkswagen); d) las plataformas como Netflix o HBO Max que, a través de suscripciones, permiten acceder a bienes y servicios digitales, y e) aquellas plataformas como Airbnb, que obtienen beneficio a través del ofrecimiento de un servicio por medio de los activos externos de los propietarios. Esta clasificación puede complementarse con la que Sequera

(2024) ha realizado recientemente, en la que especifica y aplica la propuesta de Srnicek en el ámbito urbano y, refiriéndose a plataformas urbanas, señala que existen cuatro: a) espaciales, b) cronotópicas, c) paisajísticas y d) interaccionistas de la esfera pública/privada.

Ambas clasificaciones del «capitalismo de plataforma» comparten varios elementos importantes. En primer lugar, señalan su capilaridad, que debe asombrarnos por su versatilidad y capacidad de agencia. Así, este ha sido capaz de desencadenar una *plataformización* de la vida urbana a partir de la reestructuración de los órdenes socioespaciales digitales existentes: ha exacerbado las desigualdades en las ciudades (Elwood, 2020), ha desarrollado e implementado en nuestra vida cotidiana artefactos tecnosociales que han adquirido agencia central en nuestras actividades y ha cambiado el aspecto del conjunto de nuestras relaciones urbanas. En este sentido, plataformas como Wallapop, Care.com, Uber, Deliveroo, Too Good To Go o Starlink se definen a través de una alambicada red de relaciones sociotécnicas, las cuales las constituyen como *hubs* atravesados por conexiones densamente interrelacionadas, que se incrustan en una arquitectura de prácticas extractivas, regulaciones jurídicas, interfaces digitales o redes interpersonales que facilitan una agencia distribuida entre ingenios digitales, administradores y productores de gramáticas asignificantes, así como entre actores humanos. Todos ellos operan conjuntamente en una constante reconfiguración calidoscópica de las redes y ponen en acción una dinámica relacional de poder que sobrevuela las demarcaciones convencionales entre lo social y lo tecnológico.

En segundo lugar, el capitalismo de plataforma abre la puerta a un nuevo proceso de precarización general de la ciudadanía. Por ejemplo, las denominadas «cocinas fantasmas» (*dark kitchens*), asociadas a diferentes tipos de plataformas como Deliveroo o Glovo, son establecimientos de preparación de alimentos diseñados únicamente para el servicio de entrega a domicilio y operan sin un espacio físico declarado. En ellas, los operadores de plataformas se definen como corporaciones tecnológicas, que se organizan en las ciudades con la colaboración de proveedores locales y se articulan gracias a un ciudadano que se ha convertido en *prosumidor*. Y este es precisamente el tercer rasgo que comparten ambas clasificaciones: los consumidores o usuarios de plataformas generan siempre en el mismo uso datos que alimentan inmaterialmente la capitalización de estas corporaciones, por lo que se pueden considerar al mismo tiempo productores y consumidores.

«Urbanismo inteligente» y «capitalismo de plataforma» son dos narrativas actuales que hablan de nuestras urbes; señalan, por supuesto, la condición tecno-social que han adquirido, su virtualización, la llegada de la IA y los algoritmos a su actividad diaria y, mucho más allá de todo esto, la aparición de un nuevo tipo de relaciones sociales, esto es, la «sociedad del rendimiento».

## Sociedades de rendimiento

Michel Foucault (1994) conceptualizó metodológicamente la idea de «dispositivo» como la estructura de relaciones de saber-poder en las que se inscribe el individuo y vive, en un momento histórico determinado. Los dispositivos pueden integrar tanto elementos discursivos como no discursivos, plataformas, instituciones, memorias académicas, organizaciones, arquitecturas digitales, artefactos tecnológicos, aplicaciones, regulaciones, aceptaciones de los términos de servicio, enunciados tecnocientíficos, organizaciones no gubernamentales o mecanismos de comprobación, entre otros. Un dispositivo es una especie de lógica o hilo conductor entre los diferentes elementos mencionados que el analista debe acertar a formular en su indagación académica. En nuestra opinión, la *plataformización* de la vida urbana puede ser considerada como un dispositivo. En ella se articulan, en una totalidad con sentido, elementos como artefactos tecnológicos, prácticas individuales, corporaciones extractivas, instituciones regulatorias, grupos concernidos o espacios digitalizados.

Para un autor como Han (2014), nuestras sociedades contemporáneas están abandonando el denominado «paradigma biopolítico», que se fundamenta en el concepto de «gubernamentalidad» acuñado por Foucault (2004) y opera gestionando la vida de los individuos a partir de la intervención sobre la población en su conjunto. En su lugar, se está configurando lo que él denomina «paradigma psicopolítico», asentado en dos principios: a) busca gobernar al ciudadano a través de la gestión de su psique; y b) abre una sociedad del rendimiento, es decir, unas condiciones de vida social que atienden a un imperativo de producción constante –consciente e inconsciente– que alimenta su crecimiento personal como una optimización de la vida. Como afirma Han (2014: 7), «el yo como proyecto que cree haberse liberado de las coacciones externas y de las coerciones ajenas, se somete a coacciones internas y a coacciones propias en forma de coacción de rendimiento y optimización».

La *plataformización* es, precisamente, el dispositivo que permite que opere esta sociedad del rendimiento. En ella, la producción de información por medio de dispositivos que datifican la vida, crean simulaciones y nuevos nichos de valor es fundamental (Sadowski, 2020). La psicopolítica puede entenderse como la evolución del pensamiento de Foucault y, en ella, el poder deja de ser una deter-

**La «gubernamentalidad algorítmica» se define como «un cierto tipo de racionalidad (a)normativa o (a)política que reposa sobre la recolección, la agrupación y el análisis automatizado de cantidades masivas de datos con la finalidad de modelizar, anticipar y afectar por adelantado los comportamientos posibles».**

minada relación de fuerzas, entre la sujeción y la liberación, la dominación y la resistencia, para ser entendido como un gobierno o gestión de la conducta de los otros. En tal noción del poder, actúan el *Big Data* y la IA vinculando la psique como fuerza motora del capitalismo neoliberal, con artefactos cibernéticos de producción y recolección de información que alimentan el capitalismo de plataforma, alumbrando la denominada gubernamentalidad algorítmica (Rouvroy y Berns, 2013). En ese sentido, el concepto de psicopolítica es determinante para entender las plataformas y sus tecnologías del poder, dado que a través de estas se ejerce un control sin aparentemente imponer control, tan solo seducción y motivación antes que obligación. Por ejemplo: una recomendación de un libro, por parte de la plataforma Amazon, supone un control respecto a las visitas realizadas, las compras ejercidas, el importe abonado, el idioma de referencia, el método de pago, la geolocalización del comprador, la moneda utilizada, etc., pero esta plataforma no impondrá una obligación de compra, simplemente una sugerencia para dar respuesta a una posible necesidad del sujeto de rendimiento.

En la psicopolítica, la autoexplotación, el exhibicionismo digital, el *datatísmo*, la ludificación y el agotamiento psíquico ejemplifican las dinámicas de las sociedades de rendimiento, donde las tecnologías digitales –ubicuas y ludificadas– guían la conducta de los sujetos, mostrándose en formatos amigables que nos acompañan en nuestras necesidades cotidianas y, por tanto, son muy resolutivas y necesarias en nuestros desafíos personales. Por ejemplo, una aplicación de movilidad, tipo Waze o Google Maps, nos puede indicar cuál es la ruta más rápida entre dos puntos, cuál es la más económica, o la menos peligrosa y, quizás, aquella que muestra los paisajes más bellos; sin embargo, también, nos muestra la que considera más adecuada para nosotros, lo que es todo un ejercicio de «gubernamentalidad algorítmica».

La «gubernamentalidad algorítmica» se define como «un cierto tipo de racionalidad (a)normativa o (a)política que reposa sobre la recolección, la agrupación y el análisis automatizado de cantidades masivas de datos con la finalidad de modelizar, anticipar y afectar por adelantado los comportamientos posibles» (Rouvroy y Berns, 2013: 173). El término es deudor de la noción foucaultiana de gubernamentalidad y recordemos que, para este autor, tal noción puede entenderse como la intersección de la población, el poder gubernamental y el dispositivo que permite su aplicación. Es decir, las relaciones de poder se circunscriben en un espacio que no es ni «de violencia o de lucha, ni en uno de vínculos voluntarios (todos los cuales pueden ser, en el mejor de los casos, solo instrumentos de poder), sino más bien en el área del modo de acción singular, ni belicoso ni jurídico, que es el gobierno» (Rabinow y Foucault, 1988: 254).

Para Rouvroy y Berns (2013), asistimos al nacimiento de una nueva gubernamentalidad, basada en la operación de algoritmos e IA que se articula básicamente

a partir de tres momentos centrales: a) el almacenamiento de los datos, b) la correlación de estos y c) la anticipación de las acciones por medio de la elaboración de perfiles. Es decir, la gubernamentalidad algorítmica es un proceso de extracción, tratamiento y modelización de experiencias sociales por medio de artefactos tecnológicos digitales que propician la *datificación* voluntaria e involuntaria de la vida; genera procesos de modelización automatizada en tiempo real de lo social y hace uso de sistemas de tecnologías ubicuas que entrelazan las experiencias sociales con los espacios virtuales mediante una transferencia fluida de datos que permiten reconocer patrones, provocar ajustes y modelar respuestas. Esto es, crear agencia.

La gubernamentalidad algorítmica opera gracias a la *plataformización* y supone una nueva forma de racionalidad política que se centra en la gestión y regulación de conductas gracias a la vida que permiten las nuevas urbes (Caprotti y Liu, 2022). Concebir el urbanismo de plataformas como un dispositivo que despliega este tipo de gubernamentalidad es un ejercicio heurístico que, entre otras cosas, nos permitirá entender la diferencia entre las denominadas ciudades inteligentes y las posinteligentes.

**Concebir el «urbanismo de plataformas» como un dispositivo que despliega la «gubernamentalidad algorítmica» es un ejercicio heurístico que, entre otras cosas, nos permitirá entender la diferencia entre las denominadas ciudades inteligentes y las posinteligentes.**

## Ciudades inteligentes y posinteligentes

Las *smart cities*, en su concepción original, buscan satisfacer la necesidad de innovar en los servicios públicos, mejorar las infraestructuras locales y responder de forma eficaz a las políticas de austeridad mediante la tecnificación de las ciudades. Esto permite a las corporaciones locales una gobernanza *smart* alimentada por datos a tiempo real, que posibilita tomar decisiones sobre las condiciones de vida de la ciudadanía; movilidad, calidad de aire, etc. A modo de cuadro de mando, los equipos de tecnócratas municipales disponen de nuevos conocimientos que les permiten facilitar la toma de decisiones y, por consiguiente, instaurar un nuevo tipo de gobernanza de las ciudades apoyada en nuevas fuentes de información.

Simultáneamente, las corporaciones de plataforma, autodefinidas como empresas tecnológicas, en algunos casos operan en las ciudades al margen de la legalidad local. La gobernanza de estas organizaciones es supraestatal, y atiende a un modelo de organización tecnoliberal que implementa preceptos basados en el conocimiento aportado por dispositivos tecnológicos que, en multitud de ocasiones, empujan a

tomar decisiones que entran en conflicto, por ejemplo, con las normativas locales. Es bien conocida la judicialización de las organizaciones de plataforma respecto a los derechos laborales, las regulaciones locales, o la tributación (Sancho, 2023). El marco regulatorio está, por lo tanto, desbordado a menudo por la capacidad hegemónica de las plataformas que ceden parte de las decisiones estratégicas y operacionales a los sistemas sociotécnicos desarrollados a través de IA, programados básicamente para la eficiencia de recursos y la gestión de beneficios. Podría decirse que asistimos al nacimiento de un incipiente (co)gerencialismo IA, que implica que las decisiones operacionales estén delegadas en sistemas sociotécnicos, pero, a la vez, en las netamente gerenciales. «HONG KONG, 26 de agosto de 2022 /PRNewswire/ -- NetDragon Websoft Holdings Limited (“NetDragon” o “la Compañía”, código bursátil de Hong Kong: 777), líder mundial en la creación de comunidades de Internet,

**Esto da lugar a nuevas formas de agencia y nos sitúa en el umbral de las denominadas «ciudades posinteligentes», que integran la tecnología IA tanto en su *back end* como en su *front end*. Ello supone una implementación generalizada de esta tecnología y su inserción en los espacios urbanos virtuales y físicos.**

se complace en anunciar que la Sra. Tang Yu, un robot humanoide virtual impulsado por IA, ha sido nombrado director ejecutivo rotativo de su filial insignia, Fujian NetDragon Websoft Co., Ltd. El nombramiento es un paso para ser pionero en el uso de la IA para transformar la gestión corporativa y superar la eficiencia operativa, a un nuevo nivel» (PRNewswire, 2022).

La IA desempeña, de esta forma, un papel determinante en la gestión operativa y la toma de decisiones en tiempo real en estas plataformas, permitiendo que corporaciones tecnológicas con un nivel de ingresos desorbitados estén formadas por una estructura gerencial y operativa mínima. La compañía WhatsApp, fundada inicialmente por Jan Koum, dispone actualmente de más de 2.000 millones de usuarios y su estructura está formada por apenas 3.000 trabajadores.

Esta traslación hacia una gubernamentalidad algorítmica dentro de las corporaciones de plataforma impregna la arquitectura organizativa de las *smart cities*, y avanza hacia una *plataformización* total de su funcionamiento. Casos como el de la corporación Sidewalk Labs en la ciudad de Toronto ilustran muy bien estos intentos de las mencionadas corporaciones de entrar en la gobernanza total de nuestras urbes (Goodman y Powles, 2019). Así, se podría afirmar que el núcleo inteligible del capitalismo de plataforma presenta una emergente gubernamentalidad, predecesora de la gubernamentalidad algorítmica, que aspira a la conducción de las conductas mediante la *plataformización* de la *res pública*.

Esto da lugar a nuevas formas de agencia y nos sitúa en el umbral de las denominadas «ciudades posinteligentes», que integran la tecnología IA tanto en su *back end* como en su *front end*. Ello supone una implementación generalizada de esta

tecnología y su inserción en los espacios urbanos virtuales y físicos. Así, aparecen los vehículos de limpieza autónomos, drones de reparto, robots caninos diseñados para patrullar áreas urbanas con funciones de vigilancia y seguridad, contenedores de basura inteligentes, cerebros y agentes de *software* urbanos (Cugurullo *et al.*, 2023). La implementación de estas tecnologías adquiere un nuevo significado cuando los dispositivos periféricos y auxiliares de la IA central *Brain* disponen a su vez de IA, como puede ser el caso de una simple impresora, ya que les permite tomar decisiones autónomas ante las interacciones entre humano-máquina o máquina-máquina. Una intersección que transforma sustancialmente el paradigma de las anteriores ciudades inteligentes, donde la mayoría de los dispositivos del Internet de las Cosas (IoT, por sus siglas en inglés) eran sensores que reportaban datos a una unidad central de procesamiento. En este momento, los denominados AIoT, poseen incrustada una IA, cosa que les permite dar respuesta a situaciones novedosas y, a la vez, realizar un aprendizaje continuo. Ello implica capacidades de aprendizaje automático que les permite adaptarse y tomar decisiones en función de las interacciones que realizan con su entorno. El objetivo último es facilitar el desarrollo de determinadas tareas sin la supervisión humana, lo que obliga a reflexionar a los gestores urbanos sobre modelos de cogobernanza con las IA y a los científicos sociales a examinar los modos de operar de las nuevas «microfísicas del poder», mediatizadas por artefactos dotados con capacidad de agencia.

Autores como Cugurullo sostienen que todo lo mencionado anuncia el fin de las estructuras narrativas en nuestras interacciones sociales y culturales, dado que las acciones que realizan los dispositivos dotados de IA provienen de entidades que en realidad «no son entidades conscientes, como revela el ejemplo reciente de ChatGPT» (Cugurullo *et al.*, 2023: 1.175). Y el fin de la narrativa implica la destrucción de los lazos que nos conectan y vinculan como ciudadanos que construyen las historias que edifican las polis. Es decir, la urbe posinteligente se desarrolla al margen de lo humano y sus imaginarios, ya que no los necesita. Otra característica que incide especialmente en la conformación de las ciudades posinteligentes es la velocidad. La rápida sinapsis entre máquinas-humanos y máquinas-máquinas las convierte en plataformas muy veloces, dromológicas<sup>4</sup>, que no solo aprovechan la tecnología IA, sino que también la desbordan

---

4. El concepto de dromología fue desarrollado por el filósofo y urbanista galo Paul Virilio (2006). Deriva del griego *dromos*, que significa carrera o movimiento rápido, y se refiere al estudio de la velocidad y su impacto en la sociedad. Para el autor francés, la velocidad se ha transformado en un factor decisivo en la configuración de las sociedades contemporáneas. Su análisis incide en cómo la velocidad influye en las estructuras sociales, políticas y económicas; por tanto, en el urbanismo, los mercados, la percepción del tiempo y el espacio, etc.

explorando cómo la velocidad y la aceleración influyen en la organización continua de las experiencias urbanas. La inclusión de la IA convierte a la velocidad en un agente determinante y dominante en las ciudades posinteligentes, ya que afecta a cómo interiorizamos nuestra relación con el entorno urbano, desplaza la capacidad de decisión humana hacia la decisión automatizada y la gestión del presente hacia la del futuro.

## ¿Qué es la disidencia en las ciudades posinteligentes?

La ciudad siempre se ha vinculado a la disidencia (Harvey, 2013). La urbe es fuente de poder y, por supuesto, de resistencia. Recordemos que para Foucault (1976), el poder no es algo que se posee de forma estática, sino que es una serie de relaciones dinámicas que se establecen entre individuos, instituciones y prácticas sociales, las cuales son intrincadas y se manifiestan en todos los niveles de la sociedad (Del Valle Orellana, 2014). El poder consiste, por lo tanto, en

**El poder algorítmico se vuelve opaco, e inteligible en cuanto nos acercamos a sus mecanismos de funcionamiento internos. La impenetrabilidad de los esquemas de funcionamiento asombra a legos y profesionales, por lo que no debe extrañarnos que se afirme que «el enigma de la IA generativa: hace cosas asombrosas, pero no sabemos bien por qué (...).»**

la capacidad de la articulación de un espacio o una escena, una situación (Foucault, 1976), a la que antecede una *retrosцена*, literalmente el trasfondo o escenario detrás de la escena. Este trasfondo de la corporalidad de las ciudades inteligentes son los algoritmos y su capacidad de agencia, porque «ser un agente es ser capaz de desplegar (repetidamente, en el fluir de la vida diaria) un espectro

de poderes causales, incluido el poder de influir sobre el desplegado por otros (...). Un agente deja de ser tal si pierde su aptitud de “producir una diferencia”, o sea, de ejercer alguna clase de poder» (Giddens, 1995: 51).

La *retrosцена* es la caja negra algorítmica o la IA que rige las plataformas y gobierna las ciudades inteligentes y posinteligentes. El poder algorítmico se vuelve opaco, e inteligible en cuanto nos acercamos a sus mecanismos de funcionamiento internos. La impenetrabilidad de los esquemas de funcionamiento asombra a legos y profesionales, por lo que no debe extrañarnos que se afirme que «el enigma de la IA generativa: hace cosas asombrosas, pero no sabemos bien por qué. Resolver este misterio es uno de los mayores enigmas científicos de

nuestro tiempo y un importante paso para controlar modelos futuros más potentes» (Heaven, 2024). Dicha opacidad, asentada en sofisticados mecanismos digitales, no nos debe hacer olvidar que su funcionamiento no es únicamente una cuestión mecánica de codificación que responde a proposiciones apolíticas. Dado que la ciencia y las tecnologías forman parte de esta actividad social que identificamos como política, ambas son, netamente, acciones políticas. El desarrollo de un proyecto por parte de un grupo de investigación, el equipo, sus metas, objetivos, los instrumentos digitales, las tecnologías gerenciales, los sesgos, las teorías, hipótesis, la financiación, son productos y productores de un ejercicio político (Latour, 1992).

Este ejercicio se intenta invisibilizar a través de la jerga tecnocomputacional: codificación, aprendizaje profundo (*deep learning*), scripts o centros de procesamiento de datos (CPD), ente otros conceptos. Pero se exterioriza al observar su puesta en práctica (agencia) en el entramado de usuarios y plataformas, ciudadanos y ciudades. Las decisiones que solían ser enunciadas por humanos, ahora son reescritas por algoritmos (Pasquale, 2015) y, por consiguiente, no es un error conceptual atribuirles prácticas políticas y de poder (Bucher, 2018). De este modo, se puede afirmar que las ciudades posinteligentes presentan nuevas reconfiguraciones asimétricas entre lo humano y lo no humano y generan entornos de desigualdad algorítmica, agencia y sujeción a partir del despliegue de nuevos «regímenes de la verdad», en términos foucaultianos, los cuales determinan que las proposiciones o actuaciones de los sistemas algorítmicos son infalibles respecto a cualquier otro tipo de entidad (Foucault, 2004).

Un ejemplo de lo afirmado puede ser el perfilado de personas desocupadas en Viena que muestra el servicio de empleo estatal *Arbeitsmarktservice*, que utiliza modelos poblacionales segregados y, por tanto, maximiza la exclusión individual, reproduciendo un entorno discriminatorio a partir de una recolección de datos basada en colectivos, y no en individuos (Allhutter *et al.*, 2020). Otra buena ilustración es el caso SyRI (*System Risk Indication*), un sistema algorítmico desarrollado por el Gobierno de los Países Bajos destinado a identificar posibles fraudes a la seguridad social. Su operación se centró en el análisis de ciudadanos residentes en barrios vulnerables y fue necesaria la oposición de diferentes colectivos para que las autoridades entendiesen que son los derechos humanos la base para evaluar el uso de nuevas tecnologías y no viceversa (Bekker, 2019; Van Zoonen, 2020). En ambos casos, instituciones, colectivos de diferente índole y organizaciones sociales han mostrado su rechazo ante tales despropósitos y han obtenido de sus sistemas judiciales normativas que en algunos casos han sentado jurisprudencia europea entorno a los límites de las lógicas algorítmicas, como ha pasado con SyRI.

Esta disidencia o resistencia real por parte de ciertos sectores contradice abiertamente los argumentos de Han (2014), que insiste en que los dispositivos algorítmicos no generan resistencia, sino atracción del individuo a partir de la producción de emociones positivas y su autoexplotación (ibídem, 2022). No obstante, avala la tesis foucaultiana de que allí donde existe poder, sea del tipo que fuere, también se configuran prácticas de resistencia. No olvidemos que para Foucault el poder es una relación productiva y, por tanto, si el poder es «productivo, la resistencia es inventiva» (Castro Orellana, 2017: 49), por lo que donde hay producción hay reacción de invención. O sea, es en la tríada rendimiento, resistencia, e inventiva donde la innovación aparece como el *telos* del paradigma tecnoliberal, pero también la invención se conforma como dispositivo de resistencia posbiopolítico adherido a las lógicas de las sociedades de rendimiento. Así, toda innovación tendrá siempre la posibilidad de hacer más eficaz una relación de control, pero también exhibirá la potencia de subvertir esa relación (Stiegler, 2003).

## **Bestiario de disidencias posbiopolíticas**

Michel Foucault identificó la política como aquella praxis que desarrolla conductas, actitudes, desavenencias, conflictos, resistencias, disidencias, posturas contrapuestas y tácticas que tienden a dirigir las conductas de los otros y sus discursos (Del Valle Orellana, 2014). Es decir, el ejercicio político opera a partir de tecnologías del poder que incluyen mecanismos y estrategias utilizadas por las instituciones y los individuos para controlar, regular y disciplinar a la población. Tal perspectiva biopolítica identifica la resistencia como una respuesta proactiva y productiva a las estructuras de poder dominantes en una determinada sociedad (Foucault, 1976). Para Butler, esa acción es definida siempre como colectiva y pretende revertir las tecnologías de poder mediante estrategias combativas, de confrontación y movilización «para la exigencia de justicia y para el futuro político de la libertad» (Butler y Lourties, 1998: 302). Por otro lado, la disidencia, como destaca Mouffe (2022), es un concepto que sugiere una perspectiva crítica y divergente que cuestiona los estándares establecidos sin buscar una transformación estructural inmediata. De todo esto se concluye que, mientras que la resistencia se enfrenta combativamente contra las relaciones de poder existentes, la disidencia facilita espacios de debate político, mediación y negociación democrática. En suma, estos planteamientos ofrecen un marco conceptual que diferencia la disidencia, como acción transformadora colectiva, de la resistencia como postura crítica y divergente que cuestiona las normas sociales predominantes.

Sadin (2023) ha criticado la anterior dicotomía y sostiene que ya no es operativa para describir lo que sucede con la gubernamentalidad del siglo XXI, ya que, en esta, los estados son dispositivos coactivos, pero cada vez más accesorios ante esta disposición de fuerzas. Lentamente, el gran Leviatán se diluye en una masa de microgestión controlada por algoritmos e IA, y las sociedades pierden ante esta miniaturización del poder su capacidad movilizadora. La necesidad de un rearme de lo público pasa, en opinión de este autor, por institucionalizar lo alternativo y el retorno del Estado contemporáneo a sus funciones primigenias (keynesianas). De este modo, entiende que incentivar la creación y desarrollo de diversos colectivos emancipadores con capacidad transformativa, dentro y apoyados explícitamente por las instituciones públicas, generará la única respuesta al Leviatán algorítmico. Para Sadin, la disidencia frente a la *siliconización* de la vida pasa por desprivatizar y politizar el devenir de lo social en la era de la *plataformización* (ibídem).

La reprobación o el rechazo del poder de la *plataformización* de las ciudades debe venir de la mano de nuevas alianzas entre colectivos, desafíos individuales, agrupaciones heterogéneas, mensajes insurgentes, movimientos contestatarios en defensa de los derechos humanos digitales, plataformas cooperativas, ciberactivismo, municipalismo de plataforma, reflexiones en torno a la capacidad actuación, etc., pero también a partir de los propios dispositivos de sujeción tecnoliberal como es la denominada innovación abierta (Chesbrough, 2003). Es decir, la disidencia ya no se ubica, como sucedía en la tensión clásica que hemos descrito, dentro o fuera de las relaciones que debe combatir: sino que está dentro y fuera, al mismo tiempo. No puede ser de otra manera, puesto que la gubernamentalidad algorítmica o la *plataformización* de las ciudades tampoco tiene un límite claro entre su interior y su exterior.

No resulta aventurado entender la innovación abierta como una tecnología de poder-saber, en la que las corporaciones tecnológicas (Google, Apple) se fundamentan en las comunidades a la hora de extraer una plusvalía cognitiva: «El lugar central ocupado previamente por la fuerza laboral de los trabajadores fabriles en la producción de plusvalía está siendo hoy llenado cada vez más por la fuerza laboral intelectual, inmaterial y comunicativa (...), es entonces necesario

**La reprobación o el rechazo del poder de la *plataformización* de las ciudades debe venir de la mano de nuevas alianzas entre colectivos, desafíos individuales, agrupaciones heterogéneas, mensajes insurgentes, movimientos contestatarios en defensa de los derechos humanos digitales, plataformas cooperativas, ciberactivismo, municipalismo de plataforma, reflexiones en torno a la capacidad actuación, etc., pero también a partir de los propios dispositivos de sujeción tecnoliberal como es la denominada innovación abierta.**

desarrollar una nueva teoría política del valor que pueda poner el problema de esta nueva acumulación de valor capitalista en el centro del mecanismo de explotación» (Hardt y Negri, 2000: 29).

La generación de «plusvalía cognitiva» a través de dispositivos posbiopolíticos, que apelan a la participación de comunidades, manifiesta el advenimiento de nuevos espacios de actuación comunitaria. Conectados con universidades, grupos de investigación o colectivos sociales, que permiten equipar y apropiarse a las comunidades locales de conocimientos, estrategias, artefactos digitales colaborativos y tecnologías de cooperación que facilitan contextos de resistencia y de disidencia, estos espacios pueden incidir y transformar las prácticas totalizantes del urbanismo de plataformas. Ejemplos sintomáticos pueden ser los denominados *Urban Living Labs* (Bulkeley *et al.*, 2016), la ciencia ciudadana (Blasco *et al.*, 2021) o el urbanismo experimental (Marrades *et al.*, 2021).

## Conclusiones

Conceptualizar el urbanismo de plataformas como una tecnología de poder permite analizarlo como un dispositivo que opera a partir del uso masivo de tecnologías algorítmicas. Su acción, que regula relaciones socioeconómicas y geoespaciales en las ciudades contemporáneas, se consolida a través de la matriz cultural tecnoliberal y permite el despliegue de la denominada gubernamentalidad algorítmica. En este sentido, el propósito de esta investigación ha sido desarrollar un análisis de la *plataformización* de las urbes contemporáneas desde una perspectiva psicopolítica, es decir, destacando la aparición de nuevos regímenes de gubernamentalidad que desbordan los abundantes análisis de corte biopolítico que existen.

Existe una amplia bibliografía entorno a las plataformas, sobre cómo se configuran, cómo precarizan la vida laboral, y cómo entornos productivos y culturales generan su desarrollo (Scholz, 2017; Van Doorn, 2017; Karnoven *et al.*, 2019). Sin embargo, se ha analizado menos cómo se construye este dispositivo que denominamos «plataforma» desde una perspectiva más política, no solo tecnosocial. Así, este artículo ha intentado responder a cuestiones tales como: ¿de qué está constituido este dispositivo?, ¿cómo genera agencia? Y, en su contexto, ¿qué relaciones de poder emergen?, ¿qué son las ciudades posinteligentes? o ¿qué entendemos por disidencia?

Se ha explorado cómo las plataformas se pueden entender desde diversas dimensiones políticas: como espacio físico y virtual, como una relación laboral, como un sistema de sujeción, como un régimen psicopolítico, como una entidad

dromológica, como una epistemología, como un dispositivo disciplinario, como una tecnología del yo, como un ejercicio de control de las poblaciones, como una herramienta de los estados de bienestar digitales, como una brecha en los derechos humanos, como un productor de identidades, como un concepto poscolonial, como un mecanismo de proletarización, o como un acto performativo.

En cualquier caso, resulta evidente que las plataformas operan como un mecanismo que genera poder y, a la vez, disidencia. De esa forma hay que analizarlas y caracterizarlas, a fin de comprender de manera global cómo están transformando nuestras ciudades y nuestra vida en general en estas primeras décadas del siglo XXI.

## Referencias bibliográficas

- Allhutter, Doris; Cech, Florián; Fischer, Fabián; Parrilla, Gabriel y Mager Astric. «Algorithmic profiling of job seekers in Austria: How austerity politics are made effective». *Frontiers in big data*, vol. 3, (2020). DOI: doi.org/10.3389/fdata.2020.00005 (en línea) [Fecha de consulta: 15.07.2024] <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2020.00005/full?ref=hackernoon.com>
- Bekker, Sonja. «Fundamental rights in digital welfare states: The case of SyRI in the Netherlands». *Netherlands Yearbook of International Law 2019: Yearbooks in International Law: History, Function and Future*, vol. 50, (2019), p. 289-307. DOI: doi.org/10.1007/978-94-6265-403-7\_24 (en línea) [Fecha de consulta: 15.07.2024] [https://link.springer.com/chapter/10.1007/978-94-6265-403-7\\_24#citeas](https://link.springer.com/chapter/10.1007/978-94-6265-403-7_24#citeas)
- Blasco, José Luis; Tirado, Francisco y Rovira, Joan. «Ciencia ciudadana y nuevas relaciones de poder y control». *Nómadas*, n.º 55 (2021), p. 95-109. DOI: doi.org/10.30578/nomadas.n55a6 (en línea) [Fecha de consulta: 15.07.2024] [https://editorial.ucentral.edu.co/ojs\\_uc/index.php/nomadas/article/view/3120](https://editorial.ucentral.edu.co/ojs_uc/index.php/nomadas/article/view/3120)
- Bucher, Taina. *If... then: Algorithmic power and politics*. Oxford: Oxford University Press, 2018.
- Bulkeley, Harriet; Coenen, Lars; Frantzeskaki, Niki; Hartmann, Christian; Kronsell, Annica; Mai, Lindsay; Marvin, Simon; McCormick, Kes; Van Steenberg, Frank y Voytenko Palgan, Yuliya. «Urban living labs: governing urban sustainability transitions». *Current opinion in environmental sustainability*, vol. 22, (2016), p. 13-17. DOI: <https://doi.org/10.1016/j.cosust.2017.02.003> (en línea) [Fecha de consulta: 15.07.2024] <https://www.sciencedirect.com/science/article/abs/pii/S1877343517300325>

- Butler, Judith y Marie Lourties. «Actos performativos y constitución del género: un ensayo sobre fenomenología y teoría feminista». *Debate feminista*, vol. 18, (1998), p. 296-314. [Fecha de consulta: 15.07.2024] <https://www.jstor.org/stable/42625381>
- Caprotti, Federico y Liu, Dong. «Platform urbanism and the Chinese smart city: the co-production and territorialisation of Hangzhou City Brain». *GeoJournal*, vol. 87, n.º 3 (2022), p. 1.559-1.573. DOI: [doi.org/10.1007/s10708-020-10320-2](https://doi.org/10.1007/s10708-020-10320-2) (en línea) [Fecha de consulta: 15.07.2024] <https://link.springer.com/article/10.1007/S10708-020-10320-2>
- Castro Gómez, Santiago. *Historia de la gubernamentalidad I: Razón de Estado, liberalismo y neoliberalismo en Michel Foucault*. Bogotá: Siglo del hombre, 2015.
- Castro Orellana, Rodrigo. «Foucault y la resistencia. Una gramática del concepto». *Contrastes. Revista internacional de filosofía*, vol. 22, n.º 1 (2017). DOI: <https://doi.org/10.24310/Contrastescontrastes.v22i1.3413> (en línea) [Fecha de consulta: 15.07.2024] <https://www.revistas.uma.es/index.php/contrastes/article/view/3413>
- Chesbrough, Henry. *Open innovation: a new paradigm for understanding industrial innovation*. Harvard: Harvard Business Press, 2003.
- Comisión Europea. «WiFi4EU initiative: how will it work?», (2 de junio de 2017) (en línea) [Fecha de consulta: 21.07.2024] <https://digital-strategy.ec.europa.eu/en/news/wifi4eu-initiative-how-will-it-work>
- Cortada, James W. *The digital flood: the diffusion of information technology across the U.S., Europe, and Asia*. Nueva York: Oxford University Press, 2012.
- Cugurullo, Federico; Caprotti, Federico; Cook, Matthew; Karvonen, Andrew; M Guirk, Pauline y Marvin, Simon. «The rise of AI urbanism in post-smart cities: A critical commentary on urban artificial intelligence». *Urban Studies*, vol. 61, n.º 6 (2023), p. 1.168-1.182. DOI: [doi.org/10.1177/00420980231203386](https://doi.org/10.1177/00420980231203386) (en línea) [Fecha de consulta: 15.07.2024] <https://journals.sagepub.com/doi/full/10.1177/00420980231203386>
- Del Valle Orellana, Nicolás. «Entre poder y resistencia. Tras los rastros de la política en Foucault». *Revista Enfoques: Ciencia Política y Administración Pública*, vol. 10, n.º 17 (2014), p. 147-168. DOI: [doi.org/10.60728/60d82c88](https://doi.org/10.60728/60d82c88) (en línea) [Fecha de consulta: 15.07.2024] <https://www.revistaenfoques.cl/index.php/revista-uno/article/view/69>
- Deleuze, Gilles y Guattari, Félix. *El Anti-Edipo: Capitalismo y esquizofrenia*. Buenos Aires: Editorial Paidós, 1972.
- Deng Tianhu; Keren Zhang y Zuo-Jun (Max) Shen. «A systematic review of a digital twin city: A new pattern of urban governance toward smart cities». *Journal of Management Science and Engineering*, vol. 6, n.º 2 (2021), p. 125-134. DOI: [doi.org/10.1016/j.jmse.2021.03.003](https://doi.org/10.1016/j.jmse.2021.03.003) (en línea) [Fecha de consulta: 15.07.2024] <https://www.sciencedirect.com/science/article/pii/S2096232021000238>

- Dirks, Susanne; Gurdgiev, Constantin y Keeling, Mary. «Smarter Cities for Smarter Growth: How Cities Can Optimize Their Systems for the Talent-Based Economy». *IBM Institute for Business Value* (mayo de 2010) (en línea) [Fecha de consulta: 15.07.2024] <https://ssrn.com/abstract=2001907>
- Elwood, Sarah. «Digital geographies, feminist relationality, Black and queer code studies: Thriving otherwise». *Progress in Human Geography*, vol. 45, n.º 2 (2020). DOI: [doi.org/10.1177/0309132519899733](https://doi.org/10.1177/0309132519899733) (en línea) [Fecha de consulta: 15.07.2024] <https://journals.sagepub.com/doi/abs/10.1177/0309132519899733>
- Foucault, Michel. *Histoire de la sexualité, Vol. 1: La volonté de savoir*. París: Gallimard, 1976.
- Foucault, Michel. *Dits et écrits*. París: Gallimard, 1994.
- Foucault, Michel. *Sécurité, territoire, population: Cours au Collège de France (1977-1978)*. París: Gallimard/Seuil, 2004.
- Giddens, Anthony. *La constitución de la sociedad: bases para la teoría de la estructuración*. Buenos Aires: Amorrortu, 1995.
- Goodman, Ellen P. y Powles, Julia. «Urbanism under google: lessons from sidewalk Toronto». *Fordham Law Review*, vol. 88, (2019), p. 457-498. DOI: [doi.org/10.2139/ssrn.3390610](https://doi.org/10.2139/ssrn.3390610) (en línea) [Fecha de consulta: 15.07.2024] [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3390610](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3390610)
- Han, Byung-Chul. *La sociedad de la transparencia*. Barcelona: Herder Editorial, 2013.
- Han, Byung-Chul. *Psicopolítica: Neoliberalismo y nuevas técnicas de poder*. Barcelona: Herder, 2014.
- Han, Byung-Chul. «Seis motivos por los que hoy no es posible la revolución». *El País*, (5 de noviembre de 2022) (en línea) [Fecha de consulta: 15.07.2024] <https://elpais.com/ideas/2022-11-05/seis-motivos-por-los-que-hoy-no-es-posible-la-revolucion.html>
- Hardt, Michael y Negri, Antonio. *Imperio*. Cambridge: Harvard University Press, 2000.
- Harvey, David. *Ciudades rebeldes: Del derecho de la ciudad a la revolución urbana*. Madrid: Akal, 2013.
- Heaven, Will Douglas. «El enigma de la IA generativa: hace cosas asombrosas, pero no sabemos bien por qué». *MIT Technology Review*, (8 de marzo de 2024) (en línea) [Fecha de consulta: 20.07.2024] <https://www.technologyreview.es/s/16235/el-enigma-de-la-ia-generativa-hace-cosas-asombrosas-pero-no-sabemos-bien-por-que>
- Herzberg, Caspar. *Smart Cities, Digital Nations: Building Smart Cities in Emerging Countries and Beyond*. Basingstoke: Palgrave Macmillan, 2017.
- Karnoven, Andrew; Cugurullo, Federico y Caprotti, Federico. *Inside Smart Cities*. Londres: Routledge, 2019.

- Kashef, Mohamad; Visvizi, Anna y Troisi, Orlando. «Smart city as a smart service system: Human-computer interaction and smart city surveillance systems». *Computers in Human Behavior*, vol. 124, (2021). DOI: doi.org/10.1016/j.chb.2021.106923 (en línea) [Fecha de consulta: 15.07.2024] <https://www.sciencedirect.com/science/article/pii/S0747563221002466>
- Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequence*. Londres: SAGE, 2014.
- Latour, Bruno. *Ciencia en acción: cómo seguir a los científicos e ingenieros a través de la sociedad*. Barcelona: Labor, 1992.
- Marrades, Ramón; Collin, Philippa; Catanzaro, Michelle y Mussi, Eveline. «Planning from Failure: Transforming a Waterfront through Experimentation in a Placemaking Living Lab». *Urban Planning*, vol 6, n.º 1 (2021), p. 221-234. DOI: <https://doi.org/10.17645/up.v6i1.3586> (en línea) [Fecha de consulta: 15.07.2024] <https://www.cogitatiopress.com/urbanplanning/article/view/3586>
- Mazzucato, Mariana y Collington, Rosie. *The Big Con: How the Consulting Industry Weakens Our Businesses, Infantilizes Our Governments, and Warps Our Economies*. Nueva York: Allen Lane, 2023.
- Mitchell, William James. *City of Bits. Space, Place, and the Infobahn*. Massachusetts: MIT Press, 1996.
- Morozov, Eugeny. *La locura del solucionismo tecnológico*. Buenos Aires: Katz y Capital Intelectual, 2016.
- Mouffe, Chantal. *En torno a lo político*. Buenos Aires: FCE, 2022.
- PRNewswire. «NetDragon Appoints its First Virtual CEO», (26 de agosto de 2022) (en línea) [Fecha de consulta: 25.07.2024] <https://www.prnewswire.com/news-releases/netdragon-appoints-its-first-virtual-ceo-301613062.html>
- Pasquale, Frank. *The black box society: The secret algorithms that control money and information*. Harvard: Harvard University Press, 2015.
- Rabinow, Paul y Focault, Michael. *Más allá del estructuralismo y la hermenéutica*. México: UNAM, 1988.
- Rossi, Luis Sebastián. «Ciudades, comunicación digital y pospandemia: de las smart cities al urbanismo de plataformas». *Universitas-XXI, Revista de Ciencias Sociales y Humanas*, n.º 37 (2022), p. 123-146. DOI: doi.org/10.17163/uni.n37.2022.05 (en línea) [Fecha de consulta: 15.07.2024] [http://scielo.senescyt.gob.ec/scielo.php?pid=S1390-86342022000200123&script=sci\\_arttext](http://scielo.senescyt.gob.ec/scielo.php?pid=S1390-86342022000200123&script=sci_arttext)
- Rouvroy, Antoinette y Berns, Thomas, «Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ?». *Réseaux*, n.º 177 (2013), p. 163-196. DOI: 10.3917/res.177.0163 (en línea) [Fecha de consulta: 15.07.2024] <https://www.cairn.info/revue-reseaux-2013-1-page-163.htm>

- Sadin, Eric. *Hacer disidencia: una política de nosotros mismos*. Barcelona: Herder, 2023.
- Sadowski, Jathan. *Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives, and Taking Over the World*. Cambridge: The MIT Press, 2020.
- Sancho, Jesús. «Barcelona prohibirá los supermercados fantasma». *La Vanguardia*, (13 de enero de 2023) (en línea) [Fecha de consulta: 25.07.2024] <https://www.lavanguardia.com/local/barcelona/20230113/8681208/barcelona-prohibira-todos-supermercados-fantasma-macrocinas-podran-abrir-zona-franca.html>
- Scholz, Trebor. *Uberworked and Underpaid: How Workers Are Disrupting the Digital Economy*. Cambridge: Polity Press, 2017.
- Sequera, Jorge. *La ciudad de las plataformas. Transformaciones urbanas en la era del capitalismo digital*. Barcelona: Editorial Icaria, 2024.
- Sloterdijk, Peter. *Ira y tiempo: Ensayo psicopolítico*. Madrid: Siruela, 2014.
- Srnicek, Nick. *Capitalismo de plataformas*. Buenos Aires: Caja Negra, 2017.
- Stiegler, Bernard. *La técnica y el tiempo I. El pecado de Epimeteo*. Madrid: Ediciones Cátedra, 2003.
- Townsend, Anthony M. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. Nueva York: W.W. Norton and Company, 2013.
- Van der Graaf, Shenja y Ballon, Pieter. «Navigating platform urbanism». *Technological Forecasting and Social Change*, vol. 142, (2019), p. 364-372. DOI: [doi.org/10.1016/j.techfore.2018.07.027](https://doi.org/10.1016/j.techfore.2018.07.027) (en línea) [Fecha de consulta: 15.07.2024] <https://www.sciencedirect.com/science/article/abs/pii/S0040162517318528>
- Van Doorn, Niels. «Platform labor: on the gendered and racialized exploitation of low-income service work in the ‘on-demand’ economy». *Information, Communication & Society*, vol. 20, n.º 6 (2017), p. 898-914. DOI: [doi.org/10.1080/1369118X.2017.1294194](https://doi.org/10.1080/1369118X.2017.1294194) (en línea) [Fecha de consulta: 21.07.2024] <https://www.tandfonline.com/doi/full/10.1080/1369118X.2017.1294194>
- Van Zoonen, Liesbet. «Data governance and citizen participation in the digital welfare state». *Data & Policy*, vol. 2, (2020). DOI: [doi.org/10.1017/dap.2020.10](https://doi.org/10.1017/dap.2020.10) (en línea) [Fecha de consulta: 21.07.2024] <https://www.cambridge.org/core/journals/data-and-policy/article/data-governance-and-citizen-participation-in-the-digital-welfare-state/CCF2E1914C7E2D4593D550C6BC9E4C70#article>
- Virilio, Paul. *Velocidad y política: ensayo de dromología*. Barcelona: Ediciones La Tempestad, 2006.

## Convocatoria #42: Estudios de paz y conflictos

**Editores/as invitados/as:**

Dr. Pedro Valenzuela (Pontificia Universidad Javeriana) y  
Dra. Cécile Mouly (Flacso Ecuador)

**Entrega de artículos hasta:** lunes, 6 de enero de 2025

**Publicación:** mayo de 2025

El campo de estudios de paz y conflictos ha experimentado una evolución significativa desde el fin de la Segunda Guerra Mundial en cuanto a los temas que aborda, así como los enfoques y metodologías empleados. En América Latina y el Caribe, en particular, han surgido una variedad de enfoques para hacer frente a las situaciones de conflicto armado y violencia que han afectado a diferentes países. El objetivo de este dossier es dar a conocer el campo a través de diversas investigaciones. Para ello, serán bienvenidos trabajos que aborden alguna de las siguientes líneas desde una pluralidad de enfoques:

conceptualización de la paz,  
construcción de paz desde la sociedad civil,  
mantenimiento de la paz,  
procesos de paz, negociación y mediación,  
justicia transicional,  
reconciliación,  
desarme, desmovilización y reintegración de excombatientes,  
implementación de acuerdos de paz,  
prevención de conflictos violentos,  
acción no violenta/resistencia civil.

**Envío de artículos:** solo a través de la plataforma de la revista. La fecha límite de entrega será el lunes, 6 de enero de 2025. Los artículos deberán ajustarse a las directrices para autores/as. Para la selección de artículos se utiliza un sistema de evaluación por lectores pares ciegos. El mes de publicación de URVIO #42 será en mayo de 2025. Se puede contactar con la revista ([revistaurvio@flacso.edu.ec](mailto:revistaurvio@flacso.edu.ec)) para aclarar dudas y perfilar propuestas.

# Reseñas de libros

## La colonización algorítmica de la democracia

Javier Borrás

*Investigador, CIDOB*

García-Marzá, Domingo y Calvo, Patrici

**Algorithmic democracy: a critical perspective based on deliberative democracy**

Springer, 2024

257 págs.

El impacto de la inteligencia artificial (IA) en la democracia es una cuestión ampliamente explorada en el debate académico actual. El presente libro aborda esta cuestión desde la perspectiva ética. Los autores defienden que existe un modelo en auge de «democracia algorítmica» que busca sustituir a la democracia liberal. Ello, alertan, puede conducir a un modelo político que renuncie a la participación ciudadana y a un rol relevante de la sociedad civil.

La primera parte de la obra analiza el concepto de «democracia algorítmica». Los autores consideran que vivimos un momento de «democracia a la deriva», caracterizado por una desafección de la

ciudadanía y el surgimiento de alternativas a la democracia. Una de ellas es la «democracia algorítmica», donde la IA asumiría un liderazgo tecnocrático y supuestamente neutral con el objetivo de superar los actuales problemas políticos asociados a los humanos, como la corrupción, los sesgos emocionales o la incapacidad de procesar la complejidad del mundo contemporáneo.

En la «democracia algorítmica», la calidad de la democracia se mide por los resultados cuantitativos. Factores como la libertad, la participación o la responsabilidad son secundarios e incluso negativos. Su cosmovisión es la del positivismo y la fe en los datos. No es necesaria la sociedad civil, solo el algoritmo y el ciudadano individual. Los procesos en común de deliberación y participación son prescindibles, en un mundo donde el algoritmo puede agregar las preferencias individuales y proponer resultados basándose en ellas. En contraste con esta visión, los autores consideran que una democracia sin participación y sociedad civil no es una verdadera democracia. La democracia, afirman, no solo tiene un valor utilitario, sino también moral. Mediante la participación democrática el ciudadano pone en acción su autonomía, responsabilidad y libertad. Participar en democracia le dignifica en cuanto a sujeto.

La segunda parte del libro examina ámbitos donde la «colonización» de los algoritmos está reemplazando a las personas. En primer lugar, se analiza la figura del «político virtual»: algoritmos que se «han presentado» a elecciones o «asumido cargos» en países como Nueva Zelanda, Japón o Dinamarca. Estos «políticos virtuales» se presentan como la solución a los problemas de los políticos humanos actuales: en vez de incumplir promesas o ser parciales, los políticos IA serían neutrales, incorruptibles y representarían perfectamente nuestras preferencias. Sin embargo, este discurso esconde un modelo desigual que prioriza a los ciudadanos más digitalmente conectados, y que pasa por alto que los algoritmos también están cargados de sesgos y opacidad. En contraste con la idea de «político virtual», los autores tienen una visión más positiva de los modelos de «democracia aumentada» mediante «gemelos digitales», donde cada ciudadano tendría una réplica digital que le permitiera participar más directamente en los procesos democráticos.

A continuación, se aborda el auge de las «esferas públicas digitales». En la tradicional esfera pública, existe un proceso de deliberación y participación ciudadana del cual surge una opinión pública común. En cambio, en la digital no hay creación participativa en común y el resultado se basa en la pura agregación de opiniones individuales. Los ciudadanos ya no negocian y deliberan: simplemente proveen de datos al sistema. Los autores advierten de que estas «esferas públicas digitales», además, son constantemente manipuladas mediante *bots* que hacen pasar por humanas opinio-

nes que son artificiales. Si antes las fuerzas antidemocráticas buscaban controlar la opinión pública, ahora simplemente la pueden sustituir por una artificial.

La tercera parte de la obra plantea debates éticos y ofrece propuestas. En primer lugar, los autores son escépticos de que los algoritmos puedan tener un «aprendizaje moral», ya que estos tienden al funcionalismo y a maximizar los resultados: la máquina se guía por objetivos y la mejora de indicadores, pero carece de motivación emocional y racionalidad práctica. El algoritmo tiene una actitud descriptiva hacia la ética, sin la parte normativa: reproduce la moral existente, pero no nos orienta a ser mejores.

Los autores consideran que el reto algorítmico actual debe afrontarse exigiendo «explicabilidad» a los sistemas algorítmicos. Una «explicabilidad» que puede conseguirse mediante un «diseño institucional» que promueva la transparencia y permita la participación de todos los grupos afectados por esta tecnología. Esto puede conseguirse mediante la creación de «esferas públicas»: en concreto, los autores defienden que estos espacios de participación deberían crearse dentro de las propias corporaciones tecnológicas y, mediante ellos, se podría exigir una rendición de cuentas y evitar el «lavado de cara ético». Las grandes tecnológicas deben dejar de ser mastodontes con un inmenso poder y sin apenas responsabilidades sociales. Además de la exigencia legal regulatoria, debe haber una exigencia ética que, mediante la explicabilidad y participación, pueda anticiparse a los problemas que estas compañías generan a la democracia.

Aunque este libro es un análisis original de los retos del nuevo mundo algorítmico, también tiene aspectos discutibles. Los autores sobredimensionan el impacto que la IA y los algoritmos están teniendo en la sociedad actual. Los «políticos virtuales» que analizan son más operaciones de márquetin que decisores efectivos. Hay partes del libro dedicadas al *metaverso* y otros conceptos que a pesar de tener mucho eco mediático, apenas han tenido impacto real. No toda la sociedad está abrazando la IA: en la Unión Europea, por ejemplo, existe una visión escéptica hacia estas nuevas tecnologías; en el campo de la cultura, la IA ha generado un fuerte rechazo. Por otro lado, el escenario previo a la irrupción algorítmica que plantean los autores parece ligeramente romantizado. Hablan de una esfera pública, sociedad civil y opinión pública previa a la IA en la que parece que no había constantes asimetrías de poder, interferencias o manipulación por parte de grupos de interés. Quizás más matices habrían sido positivos en este sentido.

En todo caso, aunque algunas tendencias o peligros mencionados en el libro nunca lleguen a imponerse, es importante que exista un debate al respecto para estar preparados por si estos finalmente ocurrieran. En este sentido, y en la línea marcada por los mismos autores, tener este debate sobre democracia e IA ya es algo positivo en sí mismo, como una participación colectiva ante la inercia mecánica de la «colonización» algorítmica.

## **Robots como reflejos de humanidad**

Candela Ollé

*Profesora agregada de los Estudios de Ciencias de la Información y la Comunicación, Universitat Oberta de Catalunya*

Coeckelbergh, Mark

### **La ética de los robots**

Editorial Cátedra, 2024  
200 págs.

El libro se plantea como una guía para la reflexión y toma de conciencia como ciudadanos de un mundo complejo. Mark Coeckelbergh es profesor de Filosofía de los Medios de Comunicación y Tecnología en el Departamento de Filosofía de la Universidad de Viena, con una dilatada trayectoria como escritor. El autor organiza el libro en ocho capítulos: una introducción para conocer la estructura, enfoque y objetivos (cap.1); seguidamente profundiza en diferentes tipologías de robots, desde los industriales (cap. 3), del hogar (cap. 4), los focalizados en la atención sanitaria (cap. 5), los coches autónomos (cap. 6), y los drones asesinos (cap. 7). El último capítulo aborda la ética medioambiental y, para cerrar la monografía, se incluyen un glosario, la bibliografía y un índice analítico.

Se adivina en cada capítulo la vocación didáctica del autor; incita a la reflexión constante e intenta que se comprendan mejor las implicaciones derivadas, a través de detalles que partiendo de un filósofo pueden parecer lógicos, de sentido común,

pero que actualmente cada vez son más escasos: la ética es más necesaria que nunca en una sociedad enmarañada. Los capítulos empiezan captando la atención del lector con ejemplos cinematográficos o una nota manuscrita. Y las referencias bibliográficas citadas son exhaustivas, detalle que podría permitir al autor posicionarse mejor, pero este no suele hacerlo, y va avanzando el relato con un «recomiendo mayor discusión» —una de sus frases habituales—. Sería osado, sin ser experta ni filósofa, dar más respuestas que el propio autor, pero sí que me atrevo a expresar mi posicionamiento con el cumplimiento de las leyes de la robótica o leyes Asimov, datadas del 1942, en concreto la primera: «un robot no hará daño a un ser humano, ni por inacción permitirá que un ser humano sufra daño» —citada en más de una ocasión por Coeckelbergh— y con esta premisa sigo la reseña.

Destacaría de los robots industriales su papel en el trabajo del futuro, donde la automatización de profesiones lleva consigo una pérdida de empleo y la sustitución laboral, a pesar de que los humanos serán necesarios para tareas de supervisión. Según datos de 2018 del Foro Económico Mundial, «destrezas como la creatividad, la originalidad y la iniciativa, el pensamiento crítico, la persuasión y la negociación, la resiliencia, la inteligencia emocional, el liderazgo y similares, mantendrán su valor o se volverán más valiosas» (p. 41). En el mismo contexto industrial, se plantea cuál podría ser el papel de la robótica y la IA, como tecnologías que pueden resolver problemas (medioambientales, por ejemplo), aunque también cuestionando su dependencia.

De la industria pasamos al domicilio, donde encontramos robots denominados compañeros de hogar. De la variedad de aspectos que presenta el autor destacaría la triangulación entre robots, ética y edad (jóvenes en contraposición a usuarios mayores, su salud y grado de dependencia). El uso de los robots domésticos lleva a plantear: ¿Qué sucede cuando ocurre algo malo?, ¿quién es el responsable? El autor nos adentra en una sociedad vigilada, donde tener robots también es sinónimo de dar nuestros datos y reducir nuestra privacidad.

El capítulo 4 pone el foco en la telemedicina y enfermería, desde la perspectiva de los cuidados. Cita el potencial demostrable en cirugías del robot *da Vinci*, pero no pasa por alto cómo se distribuye la responsabilidad legal. Además, expresa la preocupación cuando los robots se usan solo para ahorrar dinero sin beneficio claro para el paciente. El libro plantea de nuevo un tema de gran interés en la interacción entre los robots y las personas con demencia y/o edades avanzadas: ¿Qué se considera un buen cuidado? ¿Qué cuidados queremos?

El autor exhorta a la investigación e innovación responsables, dónde la ética esté presente desde el principio del proceso de desarrollo. Además, introduce conceptos como la agencia moral y la responsabilidad distribuida entre las muchas manos implicadas en la creación. «[...] sería altamente irresponsable no hacer responsables a los humanos de las acciones y consecuencias de los artefactos más bien limitados y, a veces, graciosamente torpes, que llamados robots, tanto automatizados como no» (p. 110), afirma. La idea es que

la brecha de responsabilidad y la brecha de conocimiento sean abarcables e insiste en la posibilidad de crear regulación que exigiera que solo se pudieran construir robots y sistemas autónomos que cumplan las condiciones de responsabilidad.

Seguidamente, cuando se abordan los coches autónomos, el autor expresa que «nos deberíamos centrar en la ética de los humanos en lugar de intentar desarrollar máquinas morales» (p. 96). Veo relevante e imprescindible el hecho de estar informados para ser conscientes de las acciones y consecuencias de la tecnología usada –una afirmación que no por ser coherente y de sentido común se debe obviar–.

El foco en los robots androides llega en el capítulo 6. A pesar de que la gente sabe que los robots son máquinas (son cosas y no tienen estados mentales, ni conciencia, ni sensibilidad) interactúan con ellos como si fueran seres humanos o animales. Consecuentemente, con el tiempo, los robots se volverán *normales*, dejaremos de temerlos y los aceptaremos. El capítulo 7 trata de las armas letales autónomas militares y los grados de autonomía de estas; se plantea la brecha de responsabilidad mediante el uso de regulación y estándares –aspectos ya comentados en otras tipologías de robots–. Para terminar, el libro añade más reflexiones acerca de por qué creamos robots y por qué deberíamos construirlos. «Los robots nos ayudan a comprender lo que nos hace humanos» (p. 154), hablar de la ética de los robots abarca temas sobre lo humano, nos sirve «para distinguirnos de otras entidades» (p. 156), afirma el autor. Una de las cuestiones reiteradas también en otros libros del autor

son las medioambientales, como el cambio climático y la destrucción del ecosistema del planeta Tierra. Sin planeta ni humanos, ¿cuál es el futuro de los robots?

Coeckelbergh establece un diálogo con una ética y filosofía moral general, desde la que se pueda aprender y a la que pueda contribuir. Incide en la dependencia de los robots, en la regulación del uso, los límites y nuestro futuro como especie, donde es clave la educación. Termino con otra cita del autor: «Dejemos que el espejo robótico nos muestre cuán democráticos somos realmente, cuánto podemos y queremos ser cuando se trata del futuro de la tecnología y de la sociedad» (p. 169).

---

## **Datificación, gobernanza y poder algorítmico: impensar las IA desde América Latina**

Ronald Sáenz Leandro  
*Investigador doctoral, Internet Interdisciplinary Institute, Universitat Oberta de Catalunya*

Danesi, Cecilia.

### **El imperio de los algoritmos: IA inclusiva, ética y al servicio de la humanidad**

Editorial Galerna, 2022  
278 págs.

En *El imperio de los algoritmos: IA inclusiva, ética y al servicio de la humanidad*, Cecilia Danesi nos guía por un análisis amplio y accesible del papel de las inteligencias artificiales (IA), los algoritmos y la datificación en la configuración de las sociedades actuales. Con un estilo amigable y divulgativo, la autora explora con afán propedéutico los conceptos esenciales que ayudan a introducir a un público no especializado los debates actuales en torno los (mal)usos de la automatización en múltiples campos, como la salud, la justicia, y las políticas públicas. Al mismo tiempo, trazando una genealogía desde su origen, hasta su evolución, aplicaciones y consecuencias, este libro ofrece una perspectiva interdisciplinaria que trasciende las disciplinas del derecho y la computación, introduciendo debates éticos y sociales que rodean al actual desarrollo de sistemas de aprendizaje automático.

La obra está estructurada en capítulos temáticos, cada uno centrado en aspectos críticos de la datificación y su influencia en la actualidad. Así, el primer capítulo establece una base conceptual al explicar qué son los algoritmos, y otros desarrollos afines como el *machine learning* y el *deep learning*. En el segundo se desmitifican algunos conceptos comunes sobre los algoritmos y se describen las diferencias entre lo que llama las IA «visibles» (como las armas autónomas) y las IA «invisibles» (sistemas de decisión automática en plataformas). En este apartado, Danesi introduce problemáticas como la «autonomía paternalista» y el determinismo tecnológico, explorando cómo la automatización puede impactar en las vidas de las personas. Su enfoque en derechos humanos se enriquece con su amplio conocimiento en el campo de la regulación y las referencias a convenciones internacionales.

De manera posterior, el análisis pasa a identificar los impactos medioambientales de las IA, especialmente en el consumo energético y de recursos que implica el entrenamiento de modelos de aprendizaje automático (*machine learning*). No obstante, la autora está lejos de plantear una lectura apocalíptica o tecnopesimista, ya que a lo largo de todo el libro plantea constantemente la necesidad de poner en la palestra el discurso optimista de la innovación frente a los dilemas éticos y perniciosos de los sistemas algorítmicos. Por eso habla de los enfoques de «algoritmos verdes» y del movimiento de IA «green by design», resaltando la importancia de la sostenibilidad en el desarrollo de nuevas tecnologías, sin dejar de lado las contradicciones inherentes a la industria tecnológica.

El cuarto capítulo aborda los sesgos algorítmicos y sus tipos, usando una clasificación intuitiva que denomina «Fórmula P.E.A.» (Programador, Entrenamiento, Aprendizaje). Esta propuesta didáctica, aunada al énfasis que la autora hace en torno a la necesidad de desarrollar algoritmos explicables y la transparencia en las aplicaciones de IA que influyen en la vida pública, constituyen los puntos destacados del libro. Asimismo, en su afán en aterrizar la presentación de las discusiones, ofrece ejemplos de casos implementados en las administraciones de varios países de América Latina y otras partes del mundo.

Entrando en el tema de las redes sociales, la autora también repasa algunos hitos conocidos como los experimentos electorales realizados por Facebook en la década anterior, como el caso Cambridge Analytica o el papel político de Twitter en campañas electorales. Su análisis expone cómo el poder de las plataformas privadas afecta el discurso público y algunos derechos como la libertad de expresión o la no discriminación. En este sentido, Danesi es clara en afirmar que no todo puede automatizarse, especialmente cuando esto afecta derechos fundamentales, resaltando la necesidad actual del desarrollo cada vez más estandarizado de auditorías algorítmicas y la naturalización de los enfoques éticos en varias etapas del desarrollo de sistemas de IA.

El sexto capítulo examina la relación entre las IA y la vigilancia, discutiendo tecnologías desde el reconocimiento facial hasta los modelos de créditos sociales. Aquí

también la autora ofrece un análisis crítico del uso de estas tecnologías en América Latina, donde la falta de transparencia, el poder de las compañías *Big Tech* y la falta de regulación agravan los problemas de privacidad y derechos en la región.

Finalmente, la obra cierra con una nota optimista: existen alternativas para mitigar los sesgos algorítmicos y promover una mayor diversidad en el diseño de las IA, el control humano, y la implementación de leyes específicas. Su enfoque humanista apunta a una fórmula que podríamos resumir como: Control humano + derecho de información = derecho a decisiones informadas. Como adenda, el apéndice del libro ofrece una excelente introducción al estado de la regulación de las IA en el ámbito iberoamericano, convirtiéndose en un recurso útil para quienes desean profundizar de forma comparada en la legislación, o bien, en los debates y enfoques regulatorios que se discuten actualmente en la región.

Con todo, la autora deja sendas abiertas para seguir profundizando sobre las temáticas planteadas. Por ejemplo, sobre cómo las personas, especialmente en contextos latinoamericanos, resisten y resignifican estas tecnologías. Esto va en línea con investigaciones empíricas actuales en ciencias sociales que indagan en prácticas particulares como las «resistencias algorítmicas» (véanse los trabajos recientes de Emiliano Treré y Tiziano Bonini), o las «domesticación mutua» entre usuarios de plataformas y algoritmos (véanse los trabajos de Ignacio Siles y Edgar Gómez-Cruz).

Asimismo, se anima a profundizar en el papel de los estados latinoamericanos como intermediarios entre las grandes

corporaciones tecnológicas y los ciudadanos. Esta es una línea que autores como Paola Ricaurte, Ignacio Siles y Edgar Gómez-Cruz han problematizado desde el concepto de «gubernamentalidad algorítmica», proponiendo estudiar la manera en que los imaginarios sociotécnicos, el *ethos* de *Silicon Valley* y las dinámicas neocoloniales y de dependencia tecnológica impactan en la implementación de sistemas de IA en la región. Esta es una preocupación relevante en torno a los debates recientes sobre el «Estado exponencial» en América Latina (véanse los trabajos de Óscar Oszlak): ¿cómo los gobiernos, en su afán de modernización, también pueden contribuir a la expansión del poder algorítmico y la vigilancia ejercidos por las corporaciones tecnológicas, incluso a expensas de los derechos de su ciudadanía? En este sentido, el libro de divulgación de Cecilia Danesi cumple su función al realizar una invitación urgente a cuestionar quién tiene el control, cómo se ejerce ese poder y las formas de mitigar su alcance.

Estos apuntes ayudan a conformar una agenda latinoamericana de investigación crítica, en dónde la tarea de «impensar las inteligencias artificiales» conlleva sopesar miradas alternativas, novedosas e interdisciplinarias en torno a la apropiación tecnológica, al tiempo que incentiva la investigación empírica regional en torno a los beneficios, daños, aportes o limitaciones de la expansión de la automatización sobre los bienes públicos y los derechos humanos.

## **La «guerra contra el género» como contestación del orden internacional**

José Antonio Sanahuja  
*Catedrático, Universidad  
Complutense de Madrid*

Barbé, Esther y Martín, Magdalena  
M. (dirs.)

### **La violencia contra las mujeres en un orden internacional en transición**

Aranzadi, 2024  
379 págs.

El fuerte ascenso de la ultraderecha en las elecciones al Parlamento Europeo —están divididos, pero si fueran un solo grupo parlamentario sería el más numeroso— y el nuevo triunfo electoral de Trump en Estados Unidos han marcado la agenda política de 2024. No son hechos puntuales, sino parte de un ciclo global y de larga duración. Ese ascenso es consecuencia de una crisis estructural de la globalización y su gobernanza, pero, en términos de agencia, la actuación de esas fuerzas iliberales y de ultraderecha es también una de sus causas. En muchos lugares, como Estados Unidos, se autodenominan conservadores, o «nuevo conservadurismo», pero en su estética, retórica y prácticas —muchas veces teñidas de populismo—, se muestran como nuevos rebeldes e insurgentes que, en nombre del pueblo, la comunidad, la patria, la «verdadera» democracia y la restauración de un supuesto «orden natural», enfrentan al sistema y a «la

casta», cuestionando la globalización y el orden liberal internacional.

La insurgencia de las nuevas ultraderechas se ha centrado en seis asuntos temáticos: 1) los valores, normas e instituciones democráticas y, en particular, al Estado de derecho y el equilibrio de poderes, 2) la *lex mercatoria* referida al libre comercio y la inversión extranjera, en favor de posiciones o bien neoproteccionistas, o bien libertarias; 3) los bienes públicos globales, como el medio ambiente o la salud; 4) la igualdad de género y el reconocimiento y derechos referidos a la diversidad sexual; 5) la migración, y 6) el soberanismo, los alineamientos geopolíticos y el reconocimiento del propio orden liberal internacional, la noción misma de Occidente, y el reconocimiento e inclusión del Sur Global. Para el analista internacional, no es difícil identificar en cada uno de ellos un conjunto o *cluster* de normas e instituciones –o, por utilizar un concepto de la obra que aquí se reseña, un «régimen complejo», con mayor o menor grado de especificidad, formalización e imperatividad–.

El volumen colectivo dirigido por Esther Barbé y Magdalena M. Martín se centra en uno de esos regímenes: el que pretende erradicar la violencia contra las mujeres, aún prevalente y que hechos como la pandemia del COVID-19 han agravado.

El primer capítulo, elaborado por las dos directoras, explica cómo se ha configurado el «régimen complejo» de normas internacionales y regionales que, como parte de un «liberalismo posnacional»,

define y trata de erradicar la violencia de género, y los tres procesos con los que se explica, desde una perspectiva eminentemente social-constructivista, su erosión y cuestionamiento: a) la fragmentación normativa, por la que ese régimen pierde coherencia jurídica, aunque pueda ganar en pluralismo legal; b) la politización de esa cuestión, con la que los actores iliberales vuelven a situar en la disputa política normas ya consensuadas y establecidas, buscando la polarización, y c) la contestación normativa, por la que se impugna la validez de una norma y/o se obstaculiza su implementación. En ese marco, la obra se organiza en tres ejes: el primero es el acervo normativo de ese régimen en el derecho internacional, con especial atención a las normas de derechos humanos, derecho internacional humanitario, y derecho penal internacional. El segundo remite a las instituciones y políticas en los ámbitos universal, regional y doméstico. El tercero, a las agendas específicas de los objetivos internacionales de desarrollo, los derechos humanos, o el marco de Mujer, Paz y Seguridad.

Los 13 estudios de caso que contiene este volumen, aplicando con disciplina y coherencia ese marco analítico, logran mostrar que la contestación al «régimen complejo» de normas para la erradicación de la violencia de género se inscribe en la doble fractura o tensión entre cosmopolitismo y comunitarismo, y entre liberalismo e iliberalismo. Sobre contestación discursiva y politización cabe destacar los capítulos sobre el movimiento antigénero en Naciones Unidas (Diego

Badell), la difícil relación entre derechos sexuales y reproductivos y violencia sexual en conflictos (Esther Barbé), los tres capítulos sobre la acción de la UE, en el plano regional y en su acción exterior (Magdalena María Martín, Carmela Barcia y Lucrecia Rubio) y el rechazo de Turquía al Convenio de Estambul (Tutku Ayhan). Como casos de contestación en la implementación de las normas, se incluyen estudios sobre violencia sexual en conflictos en África (Carmen Montero) y los casos de abusos y violencia de las fuerzas de paz de Naciones Unidas en ese continente (Raquel Vanyó). La fragmentación normativa se examina en capítulos sobre los comités de derechos humanos (Carolina Jiménez), los tribunales internacionales (Isabel Lirola), y la ausencia de normas regionales en Asia (Inés Arco y Ona Stodt). Se analiza también el proceso para categorizar la opresión de las mujeres en Afganistán como «apartheid de género» (Dorothy Estrada) y los límites de la «gobernanza liberal» de la UE que ilustra la iniciativa Spotlight en Liberia (Pol Bargués y María Martín de Almagro). Un útil capítulo final hace balance de esos aportes y revisita el marco analítico y las tesis principales de la obra.

Erradicar la violencia contra las mujeres sigue siendo una tarea imperativa y urgente. Este volumen muestra que la «guerra contra el género» de distintas fuerzas iliberales y de ultraderecha están erosionando las normas y las políticas para enfrentarla en el ámbito universal y en espacios regionales, que no fueron de

fácil construcción. Como recuerda una oportuna cita de Simone de Beauvoir al inicio del texto «(...) bastará una crisis política, económica y religiosa para que los derechos de las mujeres vuelvan a ser cuestionados». Uno de los méritos de este libro es mostrar, de manera convincente, que ese ataque es parte de un proceso más amplio de transición o, más bien, de crisis en el orden internacional liberal impulsado por esas fuerzas, coordinadas en una «internacional reaccionaria», inorgánica, pero no por ello menos real, y sin embargo un tanto ausentes en este libro, en su papel de «antiemprendedores normativos» y agentes claves de ese proceso.

---

## El conocimiento y la democracia como fuente de la legitimidad

Allan David Rodríguez Aristizábal  
*Profesor de tiempo completo, Facultad de Derecho, Universidad de Medellín; investigador predoctoral en Estudios Políticos y Jurídicos, Universidad Pontificia Bolivariana (Colombia)*

Juan Camilo Restrepo Restrepo  
*Estudiante de maestría en International Human Rights Law, Lund University (Suecia)*

Peter, Fabienne  
**The grounds of political legitimacy**  
Oxford University Press, 2023  
232 págs.

¿Por qué obedecemos al Estado y cuántos son sus órdenes legítimos? Esta es una cuestión esencial presente en Occidente desde la Antigüedad, y hoy es menester preguntarse por el fundamento contemporáneo de la legitimidad estatal, debido a los retos derivados de las nuevas tecnologías, la inestabilidad del sistema político internacional y el cuestionamiento a los ideales democrático-liberales. En este contexto, la profesora Fabienne Peter, de la University of Warwick (Reino Unido) desarrolla su texto *The grounds of political legitimacy* –Los fundamentos de la legitimidad política–.

La sección 1 (compuesta por cinco capítulos) comienza definiendo la legitimidad como la autorización dada por

un grupo para que alguien tome decisiones de carácter obligatorio (que no debe entenderse como un derecho a ordenar u obligar) (p. 3 y 9-11). Por consiguiente, es posible estar en desacuerdo con tales decisiones, pero en principio se estará obligado a cumplirlas si quien las expidió gozaba de legitimidad (cap. 1). Pero ¿cuál es el origen de esa obediencia y su correlativa legitimidad? La autora responde a esta pregunta narrando tres posibles posturas que originan la legitimidad, acompañándolas de su respectiva refutación: (i) la voluntad política (*political will*), donde la participación ciudadana determina el quehacer político, sin embargo, esta puede caer en la arbitrariedad y no necesariamente lleva a tomar decisiones *correctas* (cap. 2); (ii) los hechos normativos (*normative facts*), que en abstracto determinan lo que es legítimo por medio de criterios definitivos de lo justo, lo bueno, lo ideal (usualmente de tinte moral), y de ellos se deriva la decisión política *correcta* y su legitimidad; estos tienen como refutación el problema de la accesibilidad, en tanto que no es *accesible* a todos determinar en abstracto un criterio definitivo de lo legítimo o adecuado para los casos concretos (cap. 3), y (iii) la autoridad cognitiva política (*political cognitive authority*), que surge de los conocimientos técnico-científicos sobre un tema que permiten la generación de creencias justificadas sobre lo correcto y legítimo; empero esta postura trae la dificultad de que no siempre es posible determinar epistemológicamente la mejor posición por falta de claridad entre los expertos o falta de información (cap. 4).

En vista de las falencias de cada postura, Peter propone una perspectiva híbrida, al indicar que, cuando haya suficiente autoridad cognitiva (es decir, suficientes elementos y conocimientos fácticos), debe ser esta el fundamento de la legitimidad, y solo ante la ausencia de ella se puede recurrir a la voluntad política, entendida esta como el proceso de mediación entre posturas políticas válidas. A este fundamento de la legitimidad la denomina la responsabilidad epistémica, al permitir unir conocimientos y creencias bien fundamentadas con la voluntad ciudadana (cap. 5).

La sección 2 de la obra (compuesta por cuatro capítulos) se centra en detallar cómo se pueden tomar decisiones políticas desde la concepción de la responsabilidad epistémica, explicando que se debe garantizar una deliberación política adecuadamente organizada en la que no bastan las reglas formales de participación y debate, sino donde también cobran relevancia reglas argumentativas y basadas en evidencias, de modo tal que se justifique de manera suficiente una decisión política y así poder afirmar su legitimidad (cap. 6). En ese proceso de argumentación para la búsqueda de una decisión basada en creencias debidamente justificadas es fundamental conocer el contexto político particular y demostrar no la verdad de sus afirmaciones (que es un ideal difícil de alcanzar y podría excluir posturas que sin aspiración de verdad, son importantes), sino la suficiente evidencia de estas, al igual que contar con la participación de diferentes puntos de vista bien justificados (cap. 7).

En consecuencia, la autora establece que, ante la existencia de un cuerpo robus-

to y unificado de evidencias, de experticia suficiente y de una posición claramente identificada como la correcta, entonces debe darse prelación a la autoridad política cognitiva por encima de la voluntad política, sin por ello caer en un sistema autoritario, puesto que en el proceso político sigue existiendo deliberación, tolerancia al disenso y responsabilidad pública (cap. 8). Sin embargo, para poder llegar a dicha situación se requiere de unas condiciones particulares de claridad teórico-argumentativa y de deliberación adecuada que no siempre se evidencian en la vida política, por lo que Peter finaliza su libro explicando cómo es posible mediar los desacuerdos públicos ante la ausencia de autoridad epistémica, indicando que la mutua responsabilidad de establecer criterios de convergencia o consensos mínimos sobre aspectos sustantivos o, en su defecto, la implementación de un procedimiento democrático —entendido este como un proceso para mediar los conflictos de posiciones contrapuestas debidamente justificadas—, permiten establecer una justificación suficiente para que la decisión adoptada sea considerada como legítima (cap. 9). Para finalizar, la autora indica que el texto pretende establecer criterios de legitimación que puedan hacer frente a las dinámicas de poslegitimidad, en donde los actores políticos ni siquiera pretenden justificar sus actuaciones o decisiones, sino que, recurriendo a diversas estrategias, logran obtener obediencia sin ofrecer un sustento fáctico-argumentativo suficiente (epílogo).

En suma, puede decirse que el texto resulta de gran relevancia al ofrecer criterios normativos para la justificación

y aceptación de las decisiones políticas contemporáneas, sin recurrir a criterios de verdad o de autoridad definitiva como fundamento de la legitimidad, al dar preponderancia a las evidencias empíricas y la experticia amparadas en la democracia, para así superar la *doxa* y los riesgos del populismo, por lo que su propuesta ofrece una nueva perspectiva que amerita ser estudiada. Empero, la propuesta de Peter termina por establecer criterios abstractos que, si bien sirven como ideal, no explican por qué de facto los discursos liberales o antidemocráticos han ganado terreno recientemente. Igualmente, sus planteamientos podrían llegar a considerar como ilegítimas aquellas posturas no occidentales o que no estén basadas en los expertos o en una visión particular de democracia, que es precisamente lo cuestionado por estos nuevos actores políticos, dándoles así la razón. Por último, el modelo híbrido de Peter no tiene en cuenta las dinámicas políticas o ideológicas en la producción del conocimiento, en tanto que el consenso científico también puede reflejar sesgos o intereses específicos.

---



# Listado de revisores 2024

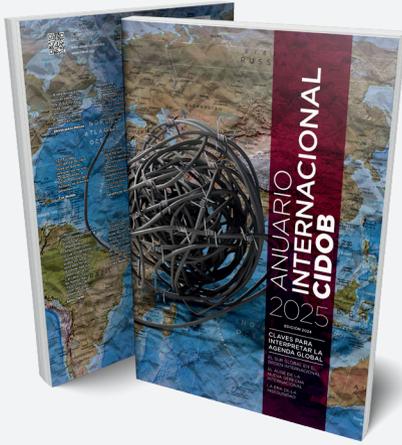
Los artículos publicados en la *Revista CIDOB d'Afers Internacionals* pasan por un riguroso proceso de evaluación externa por pares (*peer review*) de anonimato doble. Durante el año 2024 los revisores fueron los siguientes (por orden alfabético):

- 1 **Adins, Sebastien**  
Universidad Pontificia Universidad Católica del Perú (PUCP)
- 2 **Almada, Marco**  
European University Institute (EUI), Italia
- 3 **Álvarez-Verdugo, Milagros**  
Universitat de Barcelona (UB)
- 4 **Arribas, José María**  
Universidad Nacional de Educación a Distancia (UNED)
- 5 **Ayuso, Anna**  
CIDOB, Barcelona
- 6 **Bargués, Pol**  
CIDOB, Barcelona
- 7 **Barreñada Bajo, Isaías**  
Universidad Complutense de Madrid (UCM)
- 8 **Bilsky, Edgardo**  
Ciudades y Gobiernos Locales Unidos (CGLU)
- 9 **Blasi Casagran, Cristina**  
Universitat Autònoma de Barcelona (UAB)
- 10 **Cepeda, Carolina**  
Universidad Javeriana, Bogotá, Colombia
- 11 **Cervi, Laura**  
Universitat Autònoma de Barcelona (UAB)
- 12 **Cornago Prieto, Noé**  
Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)
- 13 **de Castro García, Andrés**  
Universidad Nacional de Educación a Distancia (UNED)
- 14 **de la Sierra Morón, Susana**  
Universidad de Castilla-La Mancha (UCLM)
- 15 **del Val Cid, Consuelo**  
Universidad Nacional de Educación a Distancia (UNED)
- 16 **Domínguez Mujica, Josefina**  
Universidad de Las Palmas de Gran Canaria (ULPGC)

- 17 Duch, Néstor**  
Joint Research Centre (JRC), Comisión Europea
- 18 Fernández Hellmund, Paula Daniela**  
Universidad Federal de Integración Latinoamericana (UNILA)
- 19 Fonfría Mesa, Antonio**  
Universidad Complutense de Madrid (UCM)
- 20 Foucras, Nicolas Pierre**  
Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM), México
- 21 Galceran Vercher, Marta**  
CIDOB, Barcelona
- 22 Gálvez Sánchez, Francisco Jesús**  
Universidad de Granada (UGR)
- 23 Garcés, Blanca**  
CIDOB, Barcelona
- 24 García Parra, Pio**  
Universidad Externado de Colombia
- 25 García Pérez, Rafael**  
Universidad Pablo de Olavide de Sevilla (UPO)
- 26 González Bondia, Alfonso**  
Universitat Rovira i Virgili (URV), Tarragona
- 27 González Vega, Javier A.**  
Universidad de Oviedo
- 28 Gutiérrez, Miren**  
Universidad de Deusto, Donostia
- 29 Haro, Francisco**  
Universidad de Colima, México
- 30 Hernando de Larramendi, Miguel**  
Universidad de Castilla-La Mancha (UCLM)
- 31 Jiménez, Claudia**  
Universitat Autònoma de Barcelona (UAB)
- 32 Koulu, Riikka**  
University of Helsinki, Finlandia
- 33 Lirola Delgado, Isabel**  
Universidad de Santiago de Compostela (USC)
- 34 López Giral, Dorotea**  
Universidad de Chile
- 35 Marcu, Silvia**  
Consejo Superior de Investigaciones Científicas (CSIC), Madrid
- 36 Martínez Arias, Ricardo**  
CIDOB, Barcelona

- 37 Mendiola, Ignacio**  
Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)
- 38 Molina Moreno, Valentín**  
Universidad de Granada
- 39 Negro, Samuel**  
Universidad de Navarra
- 40 Oropeza García, Héctor Arturo**  
Universidad Nacional Autónoma de México
- 41 Pasetti, Francesco**  
CIDOB, Barcelona
- 42 Perea Ozerin, Iratxe**  
Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)
- 43 Pose, Nicolás**  
Universidad de la República (Udelar), Uruguay
- 44 Rocha Pino, Manuel de Jesús**  
El Colegio de México
- 45 Rodríguez-Pérez, Adrià**  
Universitat Pompeu Fabra (UPF), Barcelona
- 46 Roldán, Javier**  
Universidad de Granada (UGR)
- 47 Romero, María Elena**  
Universidad de Colima, México
- 48 Rouvinski, Vladimir**  
Universidad Icesi, Colombia
- 49 Ruiz-Giménez Arrieta, Itziar**  
Universidad Autónoma de Madrid (UAM)
- 50 Serra i Massansalvador, Francesc**  
Universitat Autònoma de Barcelona (UAB)
- 51 Souza Ramos, Leonardo César**  
Universidade Federal de Minas Gerais, Brasil
- 52 Stuhldreher, Amalia**  
Universidad de la República (Udelar), Sede Tacuarembó, Uruguay





# ANUARIO INTERNACIONAL CIDOB

# 2025

EDICIÓN 2024

CLAVES PARA  
INTERPRETAR LA  
AGENDA GLOBAL



CONSULTA  
EL ANUARIO CIDOB  
ONLINE  
[www.anuariocidob.org](http://www.anuariocidob.org)  
[www.cidob.org](http://www.cidob.org)

## CON LAS CONTRIBUCIONES DE:

Una nueva edición del Anuario Internacional CIDOB ve la luz en un contexto internacional especialmente convulso. Este año hemos seleccionado tres temas que merecen un análisis en profundidad. El primero es la emergencia del Sur Global que, pese a su relativa indefinición en cuanto al término, se postula como entidad aglutinadora en un contrapeso a los valores y principios promovidos por Occidente. El segundo es el auge de la derecha radical en Europa y en otros muchos países y regiones del mundo. Finalmente, se aborda la denominada «era de la inseguridad», que se manifiesta tanto en el plano físico como en el emocional o psicológico, y que está alimentada por los desequilibrios de la globalización económica, las desigualdades, la revolución tecnológica o el temor ante el cambio climático, entre otros factores. El objetivo de la publicación es contribuir, desde diferentes ópticas y múltiples visiones del mundo, a una mayor comprensión de las dinámicas políticas y de las relaciones internacionales de nuestro tiempo.

### EDITA CIDOB

Elisabets, 12, 08001  
Barcelona  
[www.cidob.org](http://www.cidob.org)

### DISTRIBUYE

Edicions Bellaterra, C. de la Foneria,  
5-7, baixos. Manresa (Barcelona)  
<https://www.bellaterra.coop/es>

- Shivshankar Menon** Investigador distinguido, Centre for Social and Economic Progress (CSEP)
- Rita Abrahamsen** Profesora de Ciencias Políticas, Graduate School of Public and International Affairs, University of Ottawa
- Yun Sun** Investigadora sénior y codirectora del Programa de Asia Oriental; directora del Programa sobre China de Stimson Center
- Happymon Jacob** Profesor asociado de Diplomacia y Estudios sobre Desarme en la School of International Studies, Jawaharlal Nehru University
- Christoph Heusgen** Presidente de la Conferencia de Seguridad de Múnich
- Marina Garcés** Filosofía y activista; directora del Máster de Filosofía para los Retos Contemporáneos de la UOC
- Astra Taylor** Cineasta, escritora y activista
- Luciano Zaccara** Profesor adjunto invitado, Gulf Studies Center, Qatar University
- Cas Mudde** Polítologo y profesor de Relaciones Internacionales en la Escuela de Asuntos Públicos Internacionales, University of Georgia (Estados Unidos)
- Roberto Foa** Profesor de Ciencia Política y Políticas Públicas; Co-Director del Cambridge Centre for the Future of Democracy, Cambridge University
- Cristóbal Rovira Kaltwasser** Profesor de Ciencia Política, Pontificia Universidad Católica de Chile
- Ellen Laipson** Investigadora distinguida y presidenta emérita del Stimson Center
- Gilles Ivaldi** Investigador del Centre National de la Recherche Scientifique (CNRS) y del Centre for Political Research at Sciences Po (CEVIPOF), París
- Caroline de Gruyter** Corresponsal de asuntos europeos y columnista del periódico holandés *NRC*; columnista de *Foreign Policy* y *EUobserver*
- Jan-Werner Müller** Profesor de Ciencias Políticas, Princeton University; autor de *Democracy Rules* (Farrar, Straus and Giroux, 2021)
- Clara Mattei** Profesora asociada de Ciencias Políticas, The New School for Social Research (Nueva York)
- Hilary Cottam** Innovadora, emprendedora social y autora de *Radical Help: How we Can Remake the Relationships Between us and Revolutionise the Welfare State* (Little Brown, 2018)

**Defensa y crítica de la gobernanza algorítmica**

Daniel Innerarity

**Las *otras* geopolíticas de la IA**

Carme Colomina y Marta Galceran

**Acceso a la justicia en tiempos de IA: ¿hacia una justicia *low-cost*?**

Marta Cantero Gamito

**Opinión pública masiva: colonización algorítmica  
y sintetificación de la esfera pública**

Carlos Saura García y Patrici Calvo

**Nuevas formas de poder digital en la red:  
dimensiones de la política sexual de las *Big Tech***Águeda Gómez Suárez y Rosa M<sup>a</sup> Verdugo Matés**Los seres humanos en la toma de decisiones  
automatizada en el marco del RGPD y la Ley de IA**

Anna Levitina

**Oportunidades desde la economía social  
en la lucha contra las amenazas de la IA**

Miguel Guillén Burguillos y Eloi Serrano Robles

**La regulación europea de la IA ante los sesgos  
y riesgos de discriminación algorítmica en contextos migratorios**

Encarnación La Spina

**La llegada de la IA a las ciudades: urbanismo de plataformas  
y nuevos regímenes de gubernamentalidad**

José Luís Blasco Ejarque y Francisco Javier Tirado

La inteligencia artificial (IA) y su desarrollo hiperacelerado suponen un desafío tecnológico, económico, social y geopolítico que los grandes poderes globales abordan desde diversas visiones estratégicas, a veces contradictorias. Asistimos a un proceso de colonización algorítmica de la esfera pública y de la toma de decisiones, en el que democracia e IA se influyen mutuamente de múltiples formas. En este contexto, el número 138 de *Revista CIDOB d'Afers Internacionals* pretende contribuir a la discusión pública sobre los retos que conlleva una IA democrática. A partir de casos específicos, reflexiona sobre la gobernanza algorítmica y los riesgos que implica una IA sin cortafuegos. Se propone ofrecer una mirada propia, compleja y heterogénea al debate sobre cómo y por qué es urgente democratizar la IA.

Cuatrimestral  
de estudios  
internacionales

Barcelona,  
diciembre 2024  
ISSN 1133-6595  
ISBN: 978-84-18977-27-5  
11€

Proyecto editorial cofinanciado por la Unión Europea (UE). Las opiniones expresadas en los artículos son responsabilidad exclusiva de sus autores/as y no reflejan necesariamente los puntos de vista de la UE o la Agencia Ejecutiva Europea de Educación y Cultura (EACEA).



Cofinanciado por  
la Unión Europea

