

IIDH

Instituto Interamericano
de Derechos Humanos



Fascículo 4:

IMPACTO DE LAS REDES SOCIALES Y LA INTELIGENCIA ARTIFICIAL EN LA EROSIÓN DE LA DEMOCRACIA

Delia Ferreira Rubio

La Democracia Latinoamericana en una Encrucijada



Embajada de Noruega
Ciudad de México

© 2025 IIDH. INSTITUTO INTERAMERICANO DE DERECHOS HUMANOS

Las ideas expuestas en esta serie de fascículos son de exclusiva responsabilidad de las personas autoras y no corresponden con las ideas del IIDH/CAPEL y de sus donantes.

Ferreira Rubio, Delia M.

Impacto de las redes sociales y la Inteligencia Artificial en la erosión de la Democracia / Delia M. Ferreira Rubio --
San José, C.R. : IIDH, 2025

18 p. : 28 x 22 cm. -- (La Democracia Latinoamericana en una Encrucijada; 4)

ISBN 978-9930-514-51-1

1. Acceso a la Información. 2. Redes de información.
3. Inteligencia Artificial. 4. Democracia. 5. Sistemas democráticos.

Se permite la reproducción total o parcial de los materiales aquí publicados, siempre y cuando no sean alterados, se asignen los créditos correspondientes y se haga llegar una copia de la publicación o reproducción al editor.

EQUIPO PRODUCTOR DE LA PUBLICACIÓN:

José Thompson Jimenez

Director Ejecutivo

Sofía Vincenzi

Ángeles Fernández

Manuel Carrillo

Coordinación Académica

Andrea Tercero

Coordinación general

Walter Meoño

Corrección de estilo y diagramación

Ernesto Vega

Portada y artes finales

Instituto Interamericano de Derechos Humanos

Apartado Postal 10.081-1000 San José, Costa Rica

Tel.: (506) 2234-0404

www.iidh.ed.cr

IMPACTO DE LAS REDES SOCIALES Y LA INTELIGENCIA ARTIFICIAL EN LA EROSIÓN DE LA DEMOCRACIA

Delia M. Ferreira Rubio

EL ESCENARIO

Los índices y reportes que evalúan la calidad de la democracia en el mundo reflejan, en los últimos años, un continuo deterioro en el funcionamiento de los sistemas democráticos. Idea Internacional elabora anualmente su reporte sobre el estado de la democracia en el mundo, que analiza cuatro áreas fundamentales de un sistema democrático: derechos, participación, representación y Estado de derecho en 173 países. En la edición 2024 del *Estado Global de la Democracia* se confirma la tendencia al deterioro de la calidad de la democracia, especialmente en las áreas de representación y derechos. Solo el 24% de los países evaluados registra alguna mejora en la democracia, mientras el 47% registra un deterioro y debilitamiento del sistema. En las Américas, 44% de los países registran retrocesos, contra un 33% en los que hay avances.

El *Índice de Democracia* de The Economist Intelligence Unit analiza la calidad de la democracia en 167 países y territorios y los clasifica en cuatro tipos de régimen: democracias plenas, democracias defectuosas, regímenes híbridos y regímenes autoritarios. En la edición 2024 solo califican como democracias plenas 25 países (6,6% de la población mundial); 46 países están en la categoría de democracias defectuosas (38,4% de la población); 36 países presentan regímenes híbridos (15,7% de la población) y 60 países son regímenes autoritarios (39,2% de la población). Desde 2008 se registra un mayor deterioro en dos categorías que hacen a la esencia de la democracia: libertades civiles y elecciones

y pluralismo. Por noveno año consecutivo, se registra una baja en la evaluación promedio para los países de América Latina.

EIU Índice de Democracia 2024 - América Latina

Democracias Plenas	Puntaje general
Costa Rica	8,29
Uruguay	8,67
Democracias Defectuosas	
Argentina	6,51
Brasil	6,49
Chile	7,83
Colombia	6,35
República Dominicana	6,62
Guyana	6,11
Jamaica	6,74
Panamá	6,84
Suriname	6,79
Trinidad y Tobago	7,09
Regímenes Híbridos	
Bolivia	4,26
Ecuador	5,24
El Salvador	4,61
Guatemala	4,55
Honduras	4,98
México	5,32
Paraguay	5,92
Perú	5,69
Regímenes Autoritarios	
Cuba	2,58
Haití	2,74
Nicaragua	2,09
Venezuela	2,25

Fuente: EIU 2024, Índice de Democracia

En estas y otras mediciones sobre la calidad de la democracia, las áreas que muestran un mayor deterioro se vinculan a la representación (desconexión entre los representantes y los representados, pérdida de credibilidad de los procesos electorales); las libertades y derechos (ataques a la libertad de expresión y de prensa, vigilancia sobre los ciudadanos, desigualdad económica) y el funcionamiento del gobierno (falta de respuesta a las demandas de la sociedad, ineficiencia, atención privilegiada a grupos de interés). Frente a esta situación, no es de extrañar que se registren altos niveles de insatisfacción y desilusión con la democracia y pérdida de confianza en las instituciones y los actores políticos (Latinobarómetro, 2024) –al menos, en los actores políticos tradicionales–, lo que ha contribuido a una “autocratización” de la política.

Múltiples factores influyen en la erosión de la democracia y son de distinta naturaleza; algunos tienen connotaciones delictivas –como la corrupción y la captura del Estado–; otros son de naturaleza cultural –como la pérdida de confianza de la ciudadanía en las instituciones–; otros son de carácter institucional –como el abuso de los poderes de emergencia y la vigilancia sobre los ciudadanos–, y también los hay de carácter económico y funcional –por ejemplo, el fracaso de las políticas públicas y la falta de respuesta y solución a los problemas de la sociedad–.

La democracia no opera en el vacío. El clima de época impacta en su funcionamiento y en su eficiencia. Vivimos en tiempos de cambios acelerados, de inseguridad, de falta de certezas, actuamos en una sociedad líquida, volátil, en la que campea la posverdad y la falta de certezas. En este escenario, no es extraño que la desconfianza se difunda en todos los ámbitos de la vida social y, por supuesto, afecte también a la política. ¿Tienen las redes sociales o las nuevas herramientas de inteligencia artificial generativa algo que ver con esta situación? ¿Existe una relación causal o meramente una correlación? Para intentar contestar esta pregunta es necesario distinguir estos dos tipos de herramientas tecnológicas.

Las redes sociales (Facebook, X, Instagram, Tik Tok, entre otras) se generalizan en la primera década del siglo XXI y alcanzan su mayor difusión e impacto con la aparición de los *smartphones*, que nos permiten

estar conectados permanentemente. El desarrollo de estas plataformas está basado en un modelo de negocios (Zuboff, 2019) que requiere captar y retener la mayor cantidad de atención de los usuarios, por el mayor tiempo posible. Esto redundó en captación gratuita y continua de datos sobre los usuarios y aumenta el rédito en la publicidad.

Otro desarrollo fue el de las herramientas algorítmicas que permitieron el análisis y detección de patrones en ingentes cantidades de información –*big data*– en forma rápida superando holgadamente la capacidad de análisis de las personas. Estas herramientas se desarrollaron a través de algoritmos de recomendación (Chayka, 2024; Hoang & Fourquet, 2024) –los que determinan, por ejemplo, qué películas nos ofrece Netflix o que música nos recomienda Spotify– supuestamente basados en nuestros gustos. Los algoritmos también eran capaces de predecir el posible comportamiento de las personas; una predicción sumamente valiosa desde el punto de vista de la segmentación de la publicidad, por ejemplo.

Con el argumento de que los algoritmos eran neutrales y no cometían errores este tipo de herramientas empezaron a utilizarse en algunos países para la asistencia en la toma de decisiones sobre políticas públicas (o incluso para delegar en estas herramientas la decisión). También se aplicaron en el área de la justicia para –por ejemplo– decidir sobre la libertad condicional de una persona. En el sector privado, estas herramientas se usan para determinar primas de contratos de seguro, acceso al crédito, y en el sector de recursos humanos de las empresas para evaluar candidatos.

Aunque la Inteligencia Artificial comenzó a desarrollarse a mediados del siglo XX, en la actualidad el nombre se ha popularizado como sinónimo de lo que se conoce como Inteligencia Artificial Generativa, a partir del lanzamiento al público de ChatGPT de la empresa Open-AI a fines de 2022. Estas herramientas permiten elaborar contenido, imágenes, voces, a partir de un pedido (*prompt*) del usuario. En la actualidad la oferta de este tipo de herramienta se ha ampliado notablemente (por ejemplo, Claude, Gemini, Co-Pilot, Grok y DeepSeek –lanzada por una empresa China).

Estas herramientas están entrenadas con toda la producción humana disponible en Internet y otras bases de datos y lo que hacen al responder una pregunta/pedido es un ejercicio estadístico para predecir qué palabra debe ir después de otra para sonar como una respuesta razonable (y sintácticamente correcta) (Suleyman, 2023). Si definimos la inteligencia como la capacidad de razonar y comprender lo que se dice, claramente estas herramientas no son inteligentes, aunque con fines comerciales el nombre ayude para crear fascinación con la herramienta. Durante años nos acostumbramos a usar “navegadores”, “buscadores”, “traductores”, “correctores” y no creímos que fueran “inteligentes”. No hay nada de creativo, ni original, en estos programas; tampoco entienden lo que escriben, son sofisticados programas de computación que, además, muchas veces incurren en alucinaciones o cometen errores con la misma calidad de lenguaje, lo que los hace aun más riesgosos.

No hay que olvidar que se trata de herramientas que pueden utilizarse con fines nobles o malintencionados; que presentan oportunidades y desafíos, muchas de ellas son de uso civil y militar (Smith, 2021). Todo depende del uso que les demos. Mucho se ha hablado del riesgo existencial como una perspectiva de dominio de estos programas sobre la humanidad. Nos concentraremos, en cambio, en los efectos concretos, prácticos y actuales del uso de estas herramientas, y los daños que pueden generar, así como las estrategias que podemos poner en marcha para asegurar un diseño, distribución y uso orientado al bien común. En particular, en este artículo, en relación con la democracia.

REDES SOCIALES Y DESINFORMACIÓN

Una democracia que funcione bien requiere de comunicación, diálogo y una red de circulación de información (Harari, 2024). Si volvemos a las principales áreas de los sistemas democráticos que han sufrido un debilitamiento o erosión, comprobaremos que ellos están inmediatamente vinculados a la información. No es de extrañar que uno de los temas más

estudiados sobre la confluencia entre democracia y tecnología sea el de **la desinformación**.

La desinformación horada los cimientos de la democracia. Sin información confiable el debate se transforma en diálogo de sordos; los diversos sectores o grupos no pueden acordar o consensuar sobre la naturaleza de los hechos; el conflicto y el agravio reemplazan el diálogo; se acentúa la polarización; y todo ello socava la confianza en las instituciones y en la sociedad, confianza que es esencial para la construcción de un proyecto común y para la salud de la democracia. La desconfianza termina generalizándose. Ya nada es creíble. La sospecha de falsedad invade todo el ecosistema comunicacional. No hay certezas. Se genera una “sociedad de la desconfianza” (Han, 2024).

Los desórdenes informativos no son un fenómeno nuevo en la política y asumen múltiples formatos (campaña sucia, propaganda engañosa, teorías conspirativas, difusión de mentiras, entre otras). Las redes sociales y las herramientas de inteligencia artificial han potenciado estos desórdenes por la sofisticación que ofrecen para la generación de noticias, imágenes, videos y audios falsos y la facilidad de amplificación y circulación que facilitan.

La información como insumo y como derecho refiere a la veracidad del contenido. La veracidad de la información es la fidelidad a los hechos sobre los que se informa. En la actualidad el ecosistema informativo –en especial en lo que se refiere a la política– se caracteriza por fenómenos como la “posverdad”, los “hechos alternativos”, el “relato” político, la “narrativa adversarial” y los contenidos “sintéticos”. Cuando la verdad, la veracidad, los hechos desaparecen, también desaparecen las certezas o los acuerdos básicos sobre la facticidad (Brill, 2024). El resultado es la destrucción de la confianza. Ya no es fácil discernir lo que es cierto y lo que es falso. En ese contexto prolifera la desinformación.

A efecto de estas reflexiones utilizaremos el término desinformación de manera genérica, para referirnos a la información errónea o engañosa, la manipulación de la información, la información deliberadamente falsa,

la información descontextualizada, las noticias inventadas. Cualquiera de estos llamados “desórdenes informativos” produce como resultado la desinformación del público. El elemento objetivo de la desinformación es la tergiversación de la realidad, de los hechos. El elemento subjetivo se refiere a la intención de quien genera o propaga la desinformación.

Entre los factores que han debilitado el componente “representación” en los análisis de IDEA destaca el deterioro en la credibilidad de las elecciones. Los desórdenes informativos han marcado los procesos electorales desde 2016 y particularmente el super-ciclo electoral 2024. América Latina no ha sido la excepción y la experiencia de Brasil resulta uno de los ejemplos más interesantes en cuanto a las respuestas institucionales frente al fenómeno.

La desinformación durante el proceso electoral puede enrarecer el clima social, afectar la integridad del proceso a través de la manipulación de los votantes, fomentar una mayor polarización, acrecentar la desconfianza, y deslegitimar el resultado de las elecciones complicando incluso la gobernabilidad y la paz social. Pero quién y con qué criterio debe determinar que estamos frente a un caso de desinformación y qué medidas se aplicarán. La respuesta a estas preguntas es clave también para la integridad del proceso, que debe ser imparcial y equitativo.

Los medios tradicionales, la prensa, la radio y la televisión han sido también vehículos para difundir desinformación. Sin embargo, las reglas del periodismo y sobre todo la exigencia de verificar la información antes de publicar, actúan como barrera de protección para garantizar la calidad de la información. La función del periodista y los editores como intermediarios resulta fundamental, aunque en la actualidad el fenómeno del “periodismo militante” ha modificado la situación.

El uso de redes sociales potencia el alcance y la velocidad de circulación y por ende la potencial repercusión de la desinformación. La capacidad de viralización del mensaje, la posibilidad del anonimato, la facilidad para compartir el mensaje y el alcance global de las redes sociales implican un cambio profundo en las reglas del juego. Por otro lado, las

redes sociales favorecen la formación de cámaras de eco que incentivan la fragmentación y aumentan la polarización y el discurso agresivo (Rose-Stockwell, 2023 y Aruguete & Calvo, 2024).

Las herramientas de inteligencia artificial generativa, como ChatGPT, Copilot, Gemini, Claude, Grok, MetaAI —entre otras— han potenciado el problema especialmente porque permiten elaborar mensajes más “creíbles” y sobre todo porque facilitan la creación de imágenes, videos y mensajes de voz que impostan de manera muy efectiva a personas en situaciones inexistentes, son los llamados “*deepfakes*” (Gehringer 2024).

Los servicios de mensajería instantánea —como WhatsApp y Telegram— ofrecen otro canal muy utilizado para sembrar, viralizar y amplificar desinformación. En estos casos los mensajes llegan a través de contactos y esto los reviste de un grado importante de confiabilidad y circulan fácilmente por esas redes de contactos personales, llegando a un público amplísimo. WhatsApp jugó un rol importante en las elecciones brasileñas de 2018, que ganó Jair Bolsonaro. Al tratarse de canales encriptados se complica cualquier intento de frenar la circulación de desinformación.

Las nuevas tecnologías han permitido también la generación de nuevos medios de comunicación, los llamados “medios sintéticos”. Estos medios simulan el formato y dinámica de los medios tradicionales, incorporando un nuevo factor de ruido en la comunicación. Los medios sintéticos invaden las redes con noticiosos de televisión inexistentes, conducidos por presentadores también artificialmente creados, que replican el estilo de los verdaderos medios tradicionales en las redes.

Durante el superciclo electoral de 2024 se documentaron numerosos casos de desinformación orientados a manipular al electorado, incluso procedentes del extranjero. América Latina no fue ajena a esta tendencia. El Observatorio de Desinformación de la Universidad Complutense de Madrid ha monitoreado los casos de desinformación en las elecciones de Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, Guatemala, México, Panamá, Perú, República Dominicana, Uruguay y Venezuela. El caso más estudiado es el de Brasil donde ya en las elecciones

presidenciales de 2018 se había detectado la proliferación de campañas de desinformación y sobre todo se había identificado a las redes de mensajería instantánea –en especial WhatsApp– como uno de los canales más activos en la circulación de desinformación (Fisher, 2022).

La literatura sobre el tema detecta una clara correlación entre las tecnologías que potencian las desinformación y erosión de las instituciones democráticas. También se ha destacado la correlación entre las redes sociales, la polarización y la fragmentación (Arthur 2022). Sin embargo, resulta más difícil descifrar si existe una relación de causalidad.

ALGORITMOS E INTELIGENCIA ARTIFICIAL

Las herramientas de inteligencia artificial están aquí para quedarse e impactan en todos los campos de la experiencia humana –para bien o para mal–. Hay desarrollos muy positivos en el área de la medicina, el diagnóstico por imágenes, la investigación básica. Hay preocupación sobre sus efectos económicos y laborales a consecuencia del reemplazo de operarios o profesionales. En materia de educación hay proyectos que utilizan estas herramientas en situaciones de crisis o emergencia, por ejemplo, para facilitar la educación de las niñas y mujeres en Afganistán; pero también desde su lanzamiento existe preocupación generalizada sobre cuáles son los efectos de estas tecnologías en el rendimiento de los estudiantes.

Los gobiernos podrían utilizar –con las garantías y límites del caso– las oportunidades que estas herramientas ofrecen en materia de comunicación, información, participación, deliberación, transparencia, políticas públicas, entre otros aspectos. Por ejemplo, el uso de algoritmos para el análisis de *big data* puede ayudar en la adopción de políticas públicas que mejor se ajusten a las demandas y necesidades de la sociedad, y asistir en la tarea de los decisores. Lo que no debemos hacer –ni aceptar– es que se delegue en las herramientas la toma de la decisión.

Estas son herramientas para facilitar y efficientizar el trabajo, no para reemplazar a los decisores.

Existe la fantasía de creer que estas herramientas son neutras, perfectas e imparciales y que podrían superar los sesgos de los decisores humanos o su discrecionalidad que pueden derivar en decisiones injustas, inapropiadas, desiguales o discriminatorios. Nada más lejos de la realidad, estas herramientas pueden reproducir y perpetuar los sesgos humanos, a través de sus diseñadores y los datos de entrenamiento utilizados. Además está demostrado que estas herramientas fantasean, incurren en alucinaciones y cometen errores. Estas herramientas son incapaces de juicios valorativos, de discernimiento, de comprensión.

Una de las áreas que más se ha deteriorado en materia de democracia, según las evaluaciones que vimos al principio, es la relacionada con los derechos y las libertades civiles. Precisamente se trata de dos áreas que han sido particularmente impactadas por los riesgos de las herramientas de inteligencia artificial. Veamos algunos casos.

Las herramientas de reconocimiento facial han sido cuestionadas por afectar el derecho a la intimidad de las personas y por provocar otros daños derivados de los errores que cometen y que pueden terminar con una persona privada de su libertad sin causa alguna. Estas herramientas son utilizadas por las autoridades en muchos países, en algunos por razones de seguridad; en otros como un mecanismo directo de control y vigilancia sobre la población, tal el caso de los regímenes autoritarios. En muchas democracias, se ha cuestionado el uso del reconocimiento facial por las fuerzas policiales y se han establecido prohibiciones o limitaciones. Otra forma de ataque a la privacidad se relaciona con las herramientas algorítmicas que detectan el estado psicológico de una persona. El problema con todas estas herramientas es el de con qué objetivo se las usa. ¿Predecir las aptitudes de un candidato a un puesto? ¿Predecir una conducta delictiva? ¿Facilitar el acceso a un lugar? ¿Determinar que la persona es un objetivo militar y hay que disparar? Todos ejemplos reales. ¿Ayudan a un humano a decidir y cuál es su grado

de influencia en la decisión de la persona encargada, o directamente deciden y ejecutan una acción?

Otro de los derechos afectados es el de igualdad. Estas herramientas algorítmicas pueden producir resultados discriminatorios en función de la raza, el género, la calidad de inmigrante de una persona, la religión, o cualquier otra condición (Buolamwini, 2024). Estos resultados pueden ser consecuencia –aun involuntaria– de la incorporación de sesgos en los procesos de diseño. En algunos países se han utilizado para determinar quiénes tienen derecho a subsidios sociales o a planes de vivienda; si en el diseño del algoritmo se incluyen calificaciones negativas para algún grupo de aspirantes, la evaluación de los candidatos incluirá ese elemento y el resultado será discriminatorio. La segunda fuente de discriminación puede provenir de los datos de entrenamiento de los programas. Si estos datos están desactualizados o contienen series históricas que reflejan tendencias culturales hoy superadas, los resultados producidos por las herramientas de inteligencia artificial reproducirán aquellos estereotipos. Es el caso, por ejemplo, de lo sucedido en Estados Unidos con los algoritmos utilizados para predecir la posibilidad de reincidencia de un recluso, en caso de que se le otorgue la libertad condicional, en los que se demostró un sesgo discriminatorio contra la población de color (Christian, 2021).

El mecanismo de entrenamiento de todas estas herramientas es el recurso a la inmensa producción de la creación humana, tomada y aprovechada con fines comerciales sin licencia ni consentimiento alguno. El derecho moral de autor y el *copyright* son otras de las víctimas y esto afecta en forma general a todas las personas que han creado, compuesto, escrito, diseñado, enseñado, publicado, grabado, filmado, etc. algo que está disponible en internet. Los juicios contra las grandes plataformas empiezan a aparecer en los tribunales.

Si las herramientas reproducen sesgos discriminatorios o los datos de entrenamiento son inapropiados, es claro que el gobierno que las use producirá políticas públicas que –en el mejor de los casos– no darán

respuestas adecuadas a la sociedad o –peor aún– generarán resultados injustos con el consecuente descontento e insatisfacción de la población.

Otra de las áreas que concentran la insatisfacción de los ciudadanos es la relacionada con el funcionamiento del gobierno, la respuesta del gobierno frente a las demandas de la sociedad. Una de las herramientas de inteligencia artificial generativa más utilizada por los gobiernos, por su aparente simplicidad, es el uso de *chatbots* para interactuar con los ciudadanos brindando información o facilitando la realización de algunos trámites. Muchos de estos *chatbots* están incluidos en las páginas web de la administración como un simpático personaje que ofrece ayuda, o directamente operan a través de WhatsApp. Se supone que si el robot es entrenado con una base de datos completa de las políticas, normas y reglamentaciones que aplican, las respuestas que brinde serán acertadas (recordemos la creencia de que las herramientas son neutras y no se equivocan). El problema es que estas herramientas pueden inventar (alucinar) y cometer errores sin que los diseñadores puedan explicar el porqué. Hay casos registrados de estos *chatbots* que han proporcionado información falsa que ha llevado a los usuarios a incumplir con la ley laboral y enfrentar demandas de sus empleados. Algunos *chatbots* utilizados por autoridades electorales han brindado información errónea a los votantes sobre sus lugares de votación o sobre los plazos para hacer algún trámite relacionado con el proceso electoral, lo que puede tener consecuencias en la participación electoral. Los ejemplos pueden multiplicarse.

Es importante recordar que estas herramientas ofrecen oportunidades, pero generan importantes riesgos. La aplicación de estas herramientas en los procesos democráticos debe realizarse de forma responsable, transparente y sujeta a los controles indispensables.

Más allá de la funcionalidad y utilidad de las herramientas hay un aspecto político para tener en cuenta cuando se analiza el impacto de la inteligencia artificial en la democracia. Las empresas tecnológicas que manejan las redes sociales y las herramientas de inteligencia artificial generativa –y que anuncian el desarrollo de inteligencia artificial general, aún más

poderosa— introducen un cambio muy significativo en la distribución del poder, no solo a nivel nacional sino también a nivel geopolítico e internacional. Se trata de una decena de empresas tecnológicas que diseñan, producen, venden, distribuyen y controlan estas herramientas y además atesoran una inmensa cantidad de información; operan más allá de los límites de los territorios nacionales; manejan presupuestos superiores a los de los países; son difíciles de controlar y tienen un poder de lobby fenomenal para influir en los gobiernos. La mayoría de ellas operan desde Estados Unidos donde la regulación es mínima o inexistente. La situación es diversa en la Unión Europea, donde el año pasado se aprobó una regulación integral sobre inteligencia artificial que aplica a las plataformas que operan en el territorio europeo, aunque se advierte este año una tendencia a demorar la aplicación efectiva de las normas.

Esta nueva forma de concentración de poder ha sido descripta como “capitalismo de vigilancia” (en el que todos nosotros somos la materia prima en el proceso económico) (Zuboff, 2019) o una forma nueva de dominación el “tecno-feudalismo” (en el que todos somos siervos del sistema) (Varoufakis, 2024). Estas mega plataformas tecnológicas concentran riqueza, conocimiento y poder de control en claro desequilibrio con el poder público. No parece un camino hacia el fortalecimiento de los valores democráticos.

¿QUÉ PODEMOS HACER AL RESPECTO?

La sociedad se enfrenta a una nueva revolución tecnológica. No es la primera vez, de modo que podemos aprender sobre algunos de los errores del pasado. En general con el argumento de no frenar la innovación y el desarrollo económico o científico, la primera reacción ha sido “esperar y ver”; “no adelantarse a regular”; “la situación se acomodará”; “ya veremos cuando se produzcan daños significativos”. En la actualidad, la preocupación y la conciencia de los riesgos es generalizada, pero todavía las respuestas son dispersas y algunas cambian al son de los resultados

electorales —como ha sucedido en Estados Unidos desde comienzo de la segunda administración Trump—.

Por el momento hay tres tipos de respuesta al problema de los riesgos generados por las redes sociales y la inteligencia artificial: a) La auto-regulación, b) la regulación (nacional e internacional), y c) la alfabetización digital. No son excluyentes, obviamente.

La auto-regulación supone que las plataformas tecnológicas dueñas de las redes sociales y que diseñan herramientas de inteligencia artificial generativa son responsables de fijar sus propios límites para que sus productos no puedan ser usados para generar efectos dañinos, tanto a nivel individual como social. Con esta estrategia las empresas tecnológicas aplicarán sus propias políticas de “control de contenido” y determinarán los estándares de “ética en el diseño”, la eliminación de sesgos, la transparencia en cuanto al origen del producto, la introducción de barreras para ciertas respuestas o contenidos y el establecimiento de mecanismos de auditoría.

¿Es razonable delegar en estas poderosas plataformas la decisión sobre lo que podemos decir, sobre la forma en que lo decimos, sobre lo que podemos leer y ver, sobre quién y cómo accederá a nuestros datos y producciones? Resulta problemático poner en manos de privados semejante poder. Otra discusión es si realmente pueden ejercer razonablemente ese poder y cómo lo harán, especialmente porque estos estándares sugeridos colisionan con el modelo de negocios de la “economía de la atención”.

Las plataformas digitales han avanzado en la fijación de algunos estándares básicos para la moderación de contenido y el diseño de herramientas de inteligencia artificial. En febrero de 2024 en la Conferencia de Seguridad de Munich, Meta, OpenAI, Google, Microsoft y otras plataformas, firmaron un compromiso para combatir el uso engañoso de la inteligencia artificial en las elecciones. Entre los compromisos asumidos figuran: el desarrollo de tecnología para detectar la información maliciosa y en especial los *deepfakes*; la incorporación de

información de contexto a los posteos, indicando si han sido originados con herramientas de inteligencia artificial (por ejemplo, la inclusión de marcas de agua digitales); la colaboración con los gobiernos y la sociedad civil para la alfabetización digital de la ciudadanía y la formación de una cultura cívica digital (chequeo de fuentes, responsabilidad al compartir y circular información, uso con criterio crítico de las herramientas de generación de contenidos, entre otros recaudos).

La Unión Europea está a la avanzada en la estrategia de la regulación de la inteligencia artificial con la aprobación en abril de 2024 de la Ley de Inteligencia Artificial. Además, el Consejo de Europa aprobó el primer tratado internacional sobre inteligencia artificial: el Convenio Marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho.

En materia de legislación el desafío es complejo. No solo en lo que hace a las discusiones de fondo sobre hasta dónde es posible avanzar sin perder el balance con otros derechos esenciales para la democracia (libertad de expresión, privacidad, autonomía). Enfrentamos, y enfrentaremos, múltiples conflictos de derechos y de intereses. Además el alcance global de estas herramientas pone en tela de juicio la efectividad y aplicabilidad de las regulaciones nacionales o locales.

Los ciudadanos también tenemos una responsabilidad en la era de la IA. La prudencia en el uso de estas herramientas, la concientización sobre su potencial de daño, el criterio para discernir y chequear la información antes de circularla, son algunas de las habilidades que debemos adquirir. Debemos asumir conciencia de que estas herramientas no amplían nuestra autonomía, sino que ejercen un control sobre lo que vemos, leemos, escuchamos, compartimos y –muy importante– sobre lo que decidimos.

La generación de cultura cívica digital debería ser una prioridad en la actualidad. La alfabetización digital debería fomentar una actitud crítica de los usuarios frente a un proceso informativo en que los intermediarios (expertos, periodismo tradicional) han perdido poder y en el que prevalece la comunicación directa en la que ya no pesa la “autoridad” de la fuente.

Sería ideal que la gente tomara conciencia y entendiera el poder de las herramientas de inteligencia artificial, y cuáles son los principios de uso responsable de las mismas.

La respuesta frente a este fenomenal desafío tecnológico es el uso responsable de la inteligencia artificial por parte de todos los actores involucrados en el nuevo ecosistema, como una condición para evitar que las herramientas de inteligencia artificial contribuyan al debilitamiento de la democracia (Susskind, 2022), o –peor aún– que nos conduzcan a una especie de “algocracia” donde los algoritmos gobiernen nuestras vidas y derechos.

Señalaría, a modo de agenda elemental, los siguientes principios que deben orientar el diseño, desarrollo, distribución y uso responsable de la inteligencia artificial:

- a) No delegar la toma de decisión en herramientas algorítmicas. Utilizarlas para ayudarnos en la toma de decisiones haciendo más rápidos y eficaces los procesos.
- b) Incorporar principios éticos en el diseño, distribución y aplicación de estas tecnologías, estableciendo garantías de su uso no malicioso o dañino.
- c) Transparentar el uso de estas herramientas cuando se utilicen para determinar derechos o puedan afectar a una persona.
- d) Transparentar la naturaleza sintética de los contenidos generados.
- e) Facilitar la explicabilidad sobre los mecanismos y criterios utilizados para producir un resultado, intentando limitar los efectos derivados del problema de la “caja negra” que hace que la forma de actuación y de respuesta de estas herramientas sea inescrutable aún para sus creadores.
- f) Aplicar mecanismos que auditen el diseño de los sistemas y la calidad de los datos de entrenamiento de las herramientas.

Me gustaría terminar esta reflexión provisoria con las palabras de Henry Kissinger: "... los humanos podemos desarrollar una reverencia por las computadoras que se acerca al misticismo. Los roles de la historia, la moralidad, la justicia y el juicio humano son inciertos en un mundo así... A medida que aumentan los riesgos, nuestra respuesta debe hacerles frente" (Kissinger, 2022).

REFERENCIAS BIBLIOGRÁFICAS

Arthur, Charles 2022, *Social Warming. How Social Media Polarizes Us All*, Oneworld.

Aruguete, Natalia & Calvo, Ernesto 2024, *Nosotros contra ellos*, Siglo XXI Editores.

Brill, Steven 2022, *The Death of Truth*, Alfred Knoff.

Buolamwini, Joy 2024, *Unmasking AI*, Random House.

Chayka, Keyle 2024, *Filterworld*, Vintage Books.

Christian, Brian 2022, *The Alignment Problem*, Atlantic Books.

Economist Intelligent Unit (EIU) 2024, Índice de democracia, <https://www.eiu.com/n/campaigns/democracy-index-2024/> (última consulta 6-5-2025)

Fisher, Max 2022, *The Chaos Machine*, Little, Brown & Co.

Gehringer, Ferdinand et al. 2024, *The Influence of Deepfakes on Elections*, KAS, <https://www.kas.de/documents/d/guest/the-influence-of-deep-fakes-on-elections>, (última consulta 14-5-25).

Han, Byung-Chul 2022, *Infocracia*, Taurus.

Harari, Yuval 2024, *Nexus*, Random House.

Hoang, Lê Neguyên & Fourquet, Jean-Lou 2024, *La Dictature des Algorithmes*, Tallandier.

Idea Internacional 2024, Estado global de la democracia, <https://www.idea.int/gsod/2024/> (última consulta 6-5-25).

Kissinger, Henry et al. 2022: *The Age of AI*, John Murray.

Latinobarómetro 2024,
<https://www.latinobarometro.org/lat.jsp?Idioma=0> (última consulta 11-5-25).

Rose-Stockwell 2023, *The Outrage Machine*, Legacy.

Smith, Brad 2021, *Tools and Weapons. The Promise and Peril of the Digital Age*, Penguin Books.

Suleyman, Mustafa 2023, *The Coming Wave*, Penguin Books.

Susskind, Jamie 2022, *The Digital Republic*, Bloomsbury Publishers.

Zuboff, Shoshana 2019, *The Age of Surveillance Capitalism*, Profile Books.

Varoufakis, Yanis 2024, *Tecnofeudalismo*, Ariel.

IMPACTO DE LAS REDES SOCIALES Y LA INTELIGENCIA ARTIFICIAL EN LA EROSIÓN DE LA DEMOCRACIA

La Democracia Latinoamericana en una Encrucijada



Delia Ferreira Rubio es Abogada (Universidad Nacional de Córdoba), Doctor en Derecho por la Universidad Complutense de Madrid y Doctor Honoris Causa por la Universidad Nacional de Córdoba.

En la actualidad es miembro del Consejo Global sobre Buen Gobierno del Foro Económico Mundial y miembro del Board de RiskHubAmericas. Integra el Consejo Consultivo de SaveDemocracy y de TrustLaw de la Fundación Thompson Reuters y es Enviada de la Alianza de Gobierno Abierto.

Se desempeñó como Presidente de Transparencia Internacional (2017-2023); Miembro del Comité Directivo de la Alianza de Gobierno Abierto (2018-2021); Co-presidente del Consejo Global Anti-Corrupción del Foro Económico Mundial (2019-2022) y miembro del Board del Pacto Global de Naciones Unidas (2020-2023).

Como consultora internacional ha trabajado en diversos países para organizaciones internacionales como IFES, PNUD, OEA, BID, IDEA, NEEDS, ERIS, CAPEL, DEMOCRACY INTERNATIONAL, COUNTERPART, UNWomen, entre otras.

Es autora de numerosas publicaciones sobre transparencia y anti-corrupción, integridad y **compliance**, sistemas electorales, financiamiento de partidos políticos, ética parlamentaria y política comparada.

IIDH

Instituto Interamericano
de Derechos Humanos



Embajada de Noruega
Ciudad de México