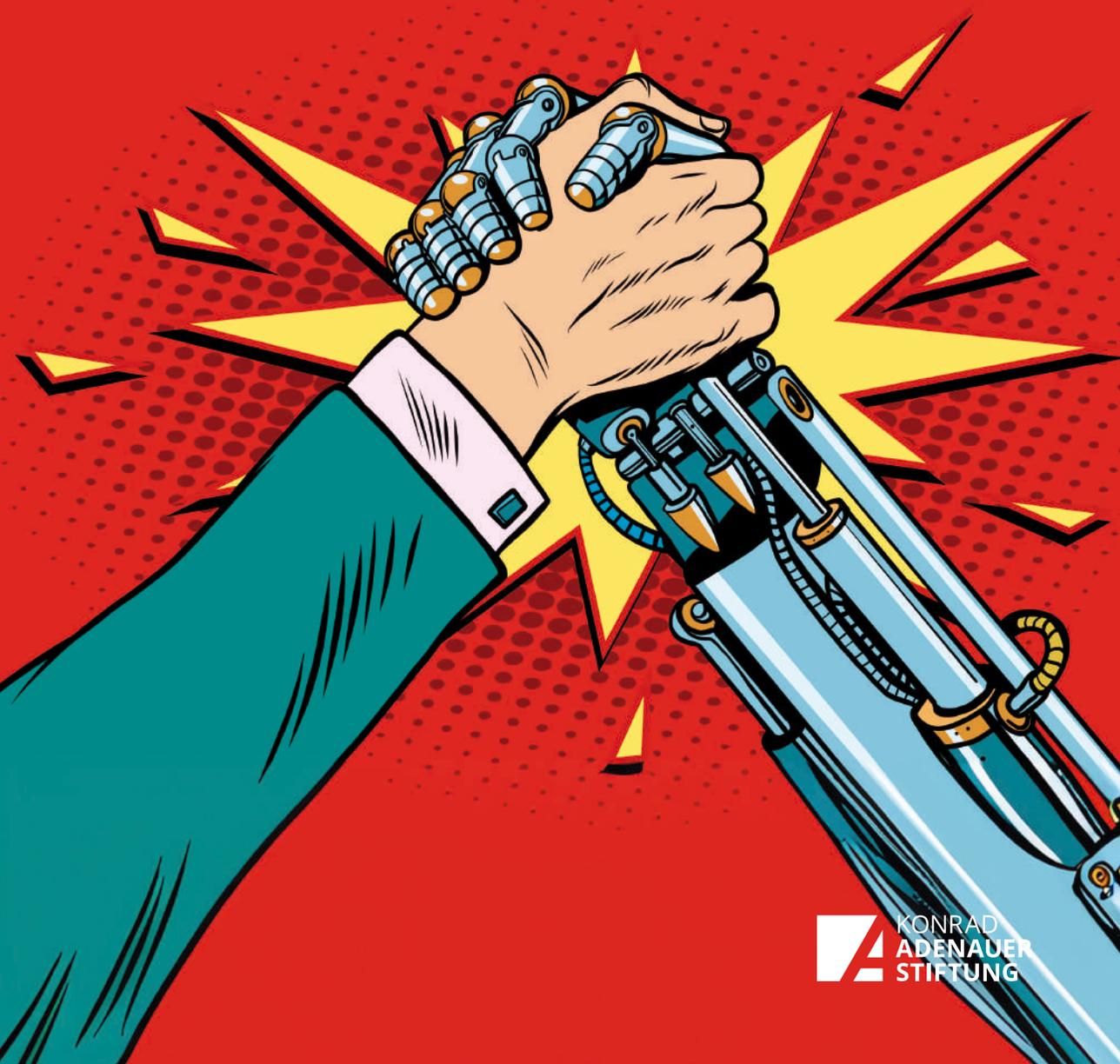


「DIÁLOGO POLÍTICO」

DEMOCRACIA ARTIFICIAL





DIÁLOGO POLÍTICO
Año XL, n.º 1, 2025

DIRECTOR
Sebastian Grundberger

SUBDIRECTOR
Jonathan Neu

COORDINACIÓN EDITORIAL
Ángel Arellano
Manfred Steffen

ASISTENTE EDITORIAL
Agustina Lombardi

CONSEJO DE REDACCIÓN
Fabiola Chambi (Bolivia)
Ana María Saavedra (Colombia)
Julio Castillo (México)
Gabriel Pastor (Uruguay)

COMMUNITY MANAGER
Yaldimar Ruiz

TRADUCCIÓN
Doris Filipovic
Marianela Santoveña
Manfred Steffen
Mirtha Tovar

CORRECCIÓN
Alejandro Coto

IMÁGENES
Shutterstock

DISEÑO Y ARMADO
Yamila Murán Leivas
@cretayam

IMPRESIÓN
Mastergraf srl
Hnos. Gil 846, Montevideo, Uruguay
www.mastergraf.com.uy

© Konrad-Adenauer-Stiftung
Plaza Independencia 749, oficina 201
11000 Montevideo, Uruguay
Tel.: +598 2902 0943

ISSN: 1688-9665

 /dialogopoliticolatinoamerica

 @dplatinooamerica

 @dplatinooamerica

 Fundación Konrad
Adenauer Montevideo

 info@dialogopolitico.org

 www.dialogopolitico.org

 www.kas.de/parteien-lateinamerika

DIÁLOGO POLÍTICO es una plataforma para el diálogo democrático entre los influenciadores políticos sobre temas de relevancia en América Latina con base en los valores de libertad, solidaridad y justicia. Conecta a la región con los grandes debates geoestratégicos en el mundo. Construye una ventana de difusión de proyectos de la Fundación Konrad Adenauer en América Latina. DIÁLOGO POLÍTICO es parte del Programa Regional Partidos Políticos y Democracia en América Latina (KAS Partidos). Tiene el objetivo de reducir la polarización política a través de un debate pluralista, constructivo e informado, orientado al bien común, para fortalecer el centro político desde sus raíces socialcristianas, liberales y conservadoras.

Los textos que se publican son de exclusiva responsabilidad de sus autores y no expresan necesariamente el pensamiento del editor. Se autoriza la reproducción total o parcial del contenido citando la fuente.

Suscríbete al boletín semanal de novedades de DIÁLOGO POLÍTICO en www.dialogopolitico.org

Índice

- 6 Prólogo
Democracia artificial
- 8 Inteligencia artificial y elecciones:
¿amenazas prematuras?
Sarah Kreps
- 22 El gobierno de los algoritmos
Daniel Innerarity
- 30 Gobernanza anticipatoria
con ayuda de la tecnología
Lydia Garrido Luzardo
- 40 ¿Es posible regular la IA?
Experiencias a nivel global
Ximena Docarmo
- 50 De las redes masivas
al voto personalizado
Jesus Delgado Valery
- 60 La lucha anticorrupción desde
el nuevo paradigma tecnológico
Denisse Rodríguez-Olivari
- 70 IA y seguridad pública en las
democracias latinoamericanas
*Gustavo Flores Macias y Bárbara
Hernández*
- 82 Lars Zimmerman: «La tecnología
no nos eximirá de las tareas de la
democracia» *Entrevista*
- 92 La libertad y la verdad secuestradas
por la inteligencia artificial
Miguel Pastorino
- 102 Personalizar las *fakes*: ¿hacia el
apocalipsis de la desinformación?
Christoph Nehring

Democracia artificial

La inteligencia artificial es omnipresente, como promesa o pesadilla. Se la asocia con las perspectivas de resolver problemas tediosos, de aprovechar mejor los recursos materiales y ahorrar tiempo. A la par, surgen preocupaciones de que sustituya a los seres humanos, ocasione desocupación masiva y liquide la creatividad y la libertad.

Muchos ven amenazada la sociedad por algoritmos que pueden potenciar, pero también manipular y sustituir a la inteligencia humana. Si definimos democracia como un debate, una conversación sobre ideas, ¿qué impacto tendrá el que nuestro interlocutor sea una máquina, un dispositivo sin sentimientos? ¿Cómo negociar sobre planes de gobierno y asignar recursos con un artefacto sin pasado ni responsabilidad respecto al futuro?

La inteligencia artificial irrumpe en tiempos signados por la amenaza de la guerra en un contexto de problemas ambientales, energéticos y civilizatorios globales. La IA se presenta como herramienta de una mejor gestión, pero exige una toma de conciencia sobre los riesgos que conlleva su aplicación sin respaldo institucional.

Desde nuestro compromiso con la democracia, en DIÁLOGO POLÍTICO queremos animar a un intercambio fundado en diversas perspectivas. Para esto invitamos a una decena de expertos internacionales.

«Para prosperar en la era digital, las democracias deben fomentar una cultura de pensamiento crítico y alfabetización mediática entre sus ciudadanos», afirma la politóloga estadounidense **Sarah Kreps**, que analiza en este número la incidencia de la IA en elecciones mediante la utilización sistemática de *deepfakes*. El director ejecutivo de Transparencia Electoral, **Jesús Delgado Valery**, agrega que el uso de la IA puede transformar las campañas mediante herramientas que permiten identificar grupos de votantes y apropiarse de sus datos personales.

El filósofo **Daniel Innerarity** duda de que «los algoritmos puedan hacerse cargo de todo el proceso político, con la eficacia que a veces prometen y con la legitimidad que debería justificar ese nuevo régimen». Para resolver esto tenemos que examinar las expectativas de la gobernanza algorítmica y sus límites.

«El poder transformador y el potencial disruptivo de la IA requieren de una gobernanza ética. La anticipación responsable y el uso alfabetizado

del futuro son elementos clave en la formulación de políticas», sostiene la experta **Lydia Garrido**.

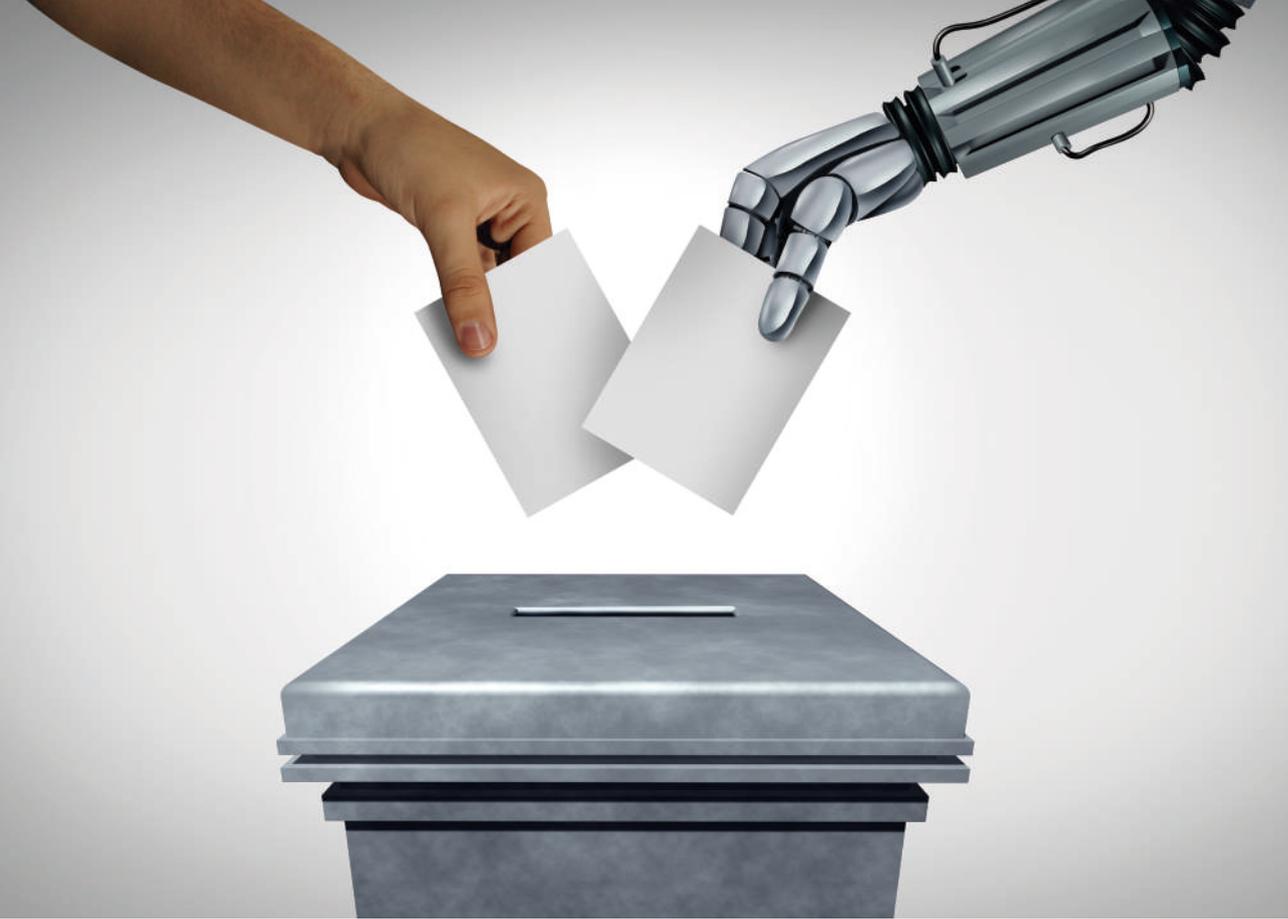
La IA es parte de una transición que es necesario regular. **Ximena Docarmo**, de InnovaLab, pone la mira sobre algunos modelos normativos posibles. Por otro lado, **Denisse Rodríguez-Olivari** expone sobre el potencial de la IA en la lucha contra la corrupción que aqueja a la institucionalidad democrática. El análisis eficiente de grandes cantidades de datos permite identificar rápidamente anomalías y actuar en consecuencia.

Otro tema que afecta a las democracias en el continente es la seguridad. **Gustavo Flores-Macías** y **Bárbara Hernández** destacan en su capítulo la necesidad del uso de la IA en forma responsable y ética para maximizar sus beneficios y mitigar los riesgos que plantea a las libertades civiles en las democracias latinoamericanas.

El filósofo **Miguel Pastorino** advierte sobre el potencial de secuestrar la verdad y la libertad de los seres humanos. En la misma línea, **Christoph Nehring**, experto del programa Disinformation, de la KAS, explora los efectos y las implicaciones de la desinformación generada por la IA y el especial rol de *influencers* y periodistas en la identificación y combate de este fenómeno.

La democracia aprende de los errores y crea un marco para dirimir diferencias y conflictos. En la transparencia en la búsqueda de soluciones, de los procesos de toma de decisión y en la rendición de cuentas se basa su legitimidad. Esto requiere de una ciudadanía informada y activa. Como dice nuestro entrevistado, el experto **Lars Zimmermann**, «La tecnología no nos eximirá de las tareas de la democracia».

Sebastian Grundberger y Manfred Steffen



Inteligencia artificial y elecciones: ¿amenazas prematuras?

SARAH E. KREPS

El nuevo paradigma tecnológico y las *deepfakes* influyen en los procesos democráticos. Los productos manipulados pueden inducir a error a las personas haciéndoles creer que determinados contenidos son reales. Las democracias deben fomentar la alfabetización mediática entre sus ciudadanos para garantizar que los valores fundacionales de la democracia perduren frente a los nuevos retos.

En 2020, uno de los principales partidos de la India, el Bharatiya Janata Party (BJP), utilizó la tecnología para crear videos de uno de sus políticos, Manoj Tiwari («Here's How Deepfakes...», 2020). Los videos mostraban a Tiwari hablando lenguas que él no habla, como el *haryanvi* y el inglés, con el objetivo de apuntar a diferentes grupos demográficos lingüísticos. El video era relativamente benigno, un intento de presentar al individuo bajo una luz positiva, pero que ponía de manifiesto las consecuencias potencialmente menos benignas de un mal uso de la tecnología. La perspectiva no era solo hipotética. Justo un año antes, en Gabón, la oficina presidencial publicó un video de su líder enfermo, Ali Bongo, que había sufrido un derrame cerebral, con el objetivo de disipar los rumores sobre su salud y la estabilidad política. El escepticismo sobre su autenticidad alimentó el malestar político y provocó un intento de golpe de Estado entre los militares que pretendían restaurar la democracia y la estabilidad en el país (Cahlan, 2020).

Con el auge de la inteligencia artificial y los crecientes ejemplos de *deepfakes* que se insinúan en el proceso democrático, los investigadores han advertido sobre casos de uso indebido más extremos. En 2023, las advertencias sobre *deepfakes* aumentaron a

medida que se acercaba 2024, un año histórico para las elecciones, con *más votantes que nunca en la historia*, en 64 países y una población del 49% del mundo (Ewe, 2023). *Newsweek* advirtió que «los *deepfakes* podrían destruir la elección de 2024» (Bickerton, 2023). En términos más generales, la proliferación de inteligencia artificial generativa como ChatGPT significa que no solo las imágenes pueden ser inauténticas, sino también el texto, como las noticias que los individuos leen sobre política (Kreps et al., 2020). Esto ha llevado a los académicos a advertir sobre los efectos de erosión de la democracia del texto generado por IA (Kreps y Kriner, 2023).

Sin embargo, las elecciones de 2024 demostraron que estas advertencias eran exageradas o, al menos, prematuras. En la mayoría de las elecciones, los *deepfakes* o los contenidos generados por IA estuvieron ausentes en gran medida, fueron vistos por relativamente pocas personas y, desde luego, no tuvieron suficiente trascendencia como para influir en las elecciones.

La pregunta entonces es ¿por qué? Con semejante potencial para destruir elecciones de formas siempre diferentes —dado que el número y el uso de *deepfakes* u otros contenidos generados por IA es casi infinito—, ¿por

qué no se ha creado IA o no ha tenido impactos consecuentes en las elecciones, consistentes con las predicciones pesimistas?

Este ensayo define en primer lugar los *deepfakes* y contenidos generados por IA, así como la razón por la que se ha predicho que la tecnología socavaría la democracia, en particular, las elecciones. A continuación, hace balance de cómo se ha utilizado la IA en diferentes elecciones de 2024, señalando la falta de impactos significativos en relación con las perspectivas teóricas. A continuación, el ensayo ofrece ideas sobre por qué la IA no ha tenido el impacto consecuente acorde con su potencial y sugiere por qué estas experiencias pasadas pueden no ser un prólogo. Concluye con reflexiones sobre posibles usos indebidos en el futuro y cómo las políticas democráticas deben permanecer vigilantes y digitalmente alfabetizadas.

¿Qué es la IA generativa y cuál es la amenaza potencial para la democracia?

La inteligencia artificial generativa es un subconjunto de la inteligencia artificial capaz de crear nuevos contenidos como texto, imágenes, audio y video (Martinau, 2023). La tecnología se basa en modelos de aprendizaje automático, en particular, los que implican aprendizaje profundo, para generar resultados que imitan los datos del mundo real. Aunque el auge de las ChatGPT dirigidas al consumidor ha provocado la proliferación de la IA generativa basada en texto y la preocupación por las alteraciones democráticas, los *deepfakes* ya han creado videos realistas pero falsos de candidatos políticos o personajes públicos y se han utilizado para desinformar.

Los *deepfakes* son medios de comunicación sintéticos en los que una persona, en una imagen o video existente, se sustituye

❖ **La preocupación fundamental con los *deepfakes* es que su verosimilitud entre la realidad y el audio, el video o las imágenes manipulados puede inducir a error a las personas haciéndoles creer que el contenido es real.** ❖

por otra similar mediante el uso de inteligencia artificial (IA) y técnicas de aprendizaje automático (Yasar, 2024). El término proviene de la combinación de *deep learning* (aprendizaje profundo), que es un término que combina aprendizaje automático, y *fake* (falso), lo que implica algo que no es auténtico. Los *deepfakes* se crean utilizando un tipo de inteligencia artificial llamada *generative adversarial networks* (GAN). Se puede pensar en las GAN como en un par de artistas digitales. Un artista, el *generador*, intenta crear imágenes o videos falsos que parezcan reales, mientras que el otro artista, el *discriminador* intenta detectar las falsificaciones. Mediante este proceso de iterativo, el *generador* mejora su capacidad de crear contenidos falsos de aspecto realista. Otro tipo de IA, llamado *autocodificadores*, ayuda aprendiendo los patrones de las imágenes o videos reales y utilizando ese conocimiento para recrear contenidos similares, pero falsos.

La preocupación fundamental con los *deepfakes* es que su verosimilitud entre la realidad y el audio, el video o las imágenes manipulados puede inducir a error a las personas haciéndoles creer que el contenido es real. Los *deepfakes* suponen teóri-

» Podrían utilizarse para crear falsas narrativas haciendo parecer que figuras políticas han dicho o hecho cosas que en realidad nunca hicieron « .

camente una amenaza para las elecciones democráticas debido a su potencial para manipular la percepción pública, difundir información errónea y socavar la confianza en el proceso electoral.

Por ejemplo, podrían utilizarse para crear falsas narrativas haciendo parecer que figuras políticas han dicho o hecho cosas que en realidad nunca hicieron. Esto puede inducir a error al público e influir en la opinión pública basándose en información fabricada. El de Gabón es un buen ejemplo, ya que ha erosionado la confianza en las figuras públicas y los medios de comunicación y ha socavado la integridad de la democracia (Cahlan, 2020).

De hecho, los estudiosos han advertido y demostrado que las personas no pueden distinguir el contenido de las noticias generado por IA del contenido real de las noticias (Kreps et al., 2020), que la propaganda generada por IA es persuasiva para las personas en formas que podrían lograr el objetivo de la manipulación psicológica (Goldstein et al., 2024) y que los miembros del Congreso no pueden distinguir entre los mensajes de los electores generados por IA y los generados por humanos. Potencialmente, esto puede afectar a la agenda legislativa, dado el potencial de escribir mensajes de promoción a escala (Kreps y Kriner, 2023). Si bien estos estudios se refieren a contenidos generados por IA en general y no específicamente a *deepfakes*, los puntos son los mismos.

Las personas no pueden distinguir entre los contenidos generados por IA y los generados por humanos, los contenidos pueden manipular a las personas y estas tecnologías tienen la capacidad de proliferar y escalar.

Aunque la propaganda y la desinformación estén disponibles desde hace mucho tiempo, la IA generativa es diferente en varios aspectos. Por ejemplo, los *deepfakes* pueden crear audio, imágenes y videos muy realistas, que son mucho más convincentes que los métodos tradicionales que tratan de tergiversar la realidad, lo que hace que sea más difícil para las personas detectar la falta de autenticidad. Asimismo, la naturaleza auditiva o visual de algunos tipos de contenidos generados puede tener un mayor impacto emocional en las personas que la desinformación basada en texto, lo que hace que el mensaje sea más memorable y persuasivo. La proliferación de la tecnología de IA generativa también hace que sea más rápido, más barato y más fácil —incluso en tiempo real— para cualquier persona producir contenidos sofisticados que antes solo era posible para actores estatales o para organizaciones con buenos recursos. Estas diferencias hacen de los *deepfakes* una herramienta especialmente potente para la desinformación y la manipulación (Verma, 2023), lo que plantea nuevos retos para mantener la integridad del discurso público y los procesos democráticos.

Más allá del potencial manipulador directo de los contenidos generados por la IA, otra preocupación sobre los *deepfakes* no son ni siquiera los *deepfakes* en sí mismos, sino la forma en que la proliferación de *deepfakes* crea dudas sobre la veracidad de la información en general. En otras palabras, si la gente llega a pensar que cualquier cosa puede ser falsa, es posible que no confíe en que algo sea cierto, lo que provocaría una erosión de las fuentes de información.

El llamado dividendo del mentiroso —en el que la información falsa está tan extendida que la negación de la realidad se hace más plausible— se convierte en un problema mayor a medida que la existencia de *deep-fakes* socava la confianza en los medios de comunicación y la información auténticos (Jackson et al., 2024).

Funcionarios gubernamentales de todo el mundo se han tomado en serio los riesgos de la IA generativa. El Departamento de Ciberseguridad e Infraestructura de Estados Unidos, por ejemplo, ha advertido de que la IA generativa «puede amplificar los riesgos existentes para la infraestructura electoral». La revista *Wired*, especializada en tecnología, llamó a 2024 «el año de la elección de la IA generativa» (Elliot, 2024). Aunque el riesgo teórico de daño existe, ya sea a través de texto o, más probablemente, a través de imágenes, ¿las experiencias han confirmado realmente las preocupaciones? La siguiente sección ofrece un cauto optimismo, sugiriendo que el uso o, más bien, el mal uso de la IA generativa ha sido limitado a pesar de la amplia disponibilidad de la tecnología.

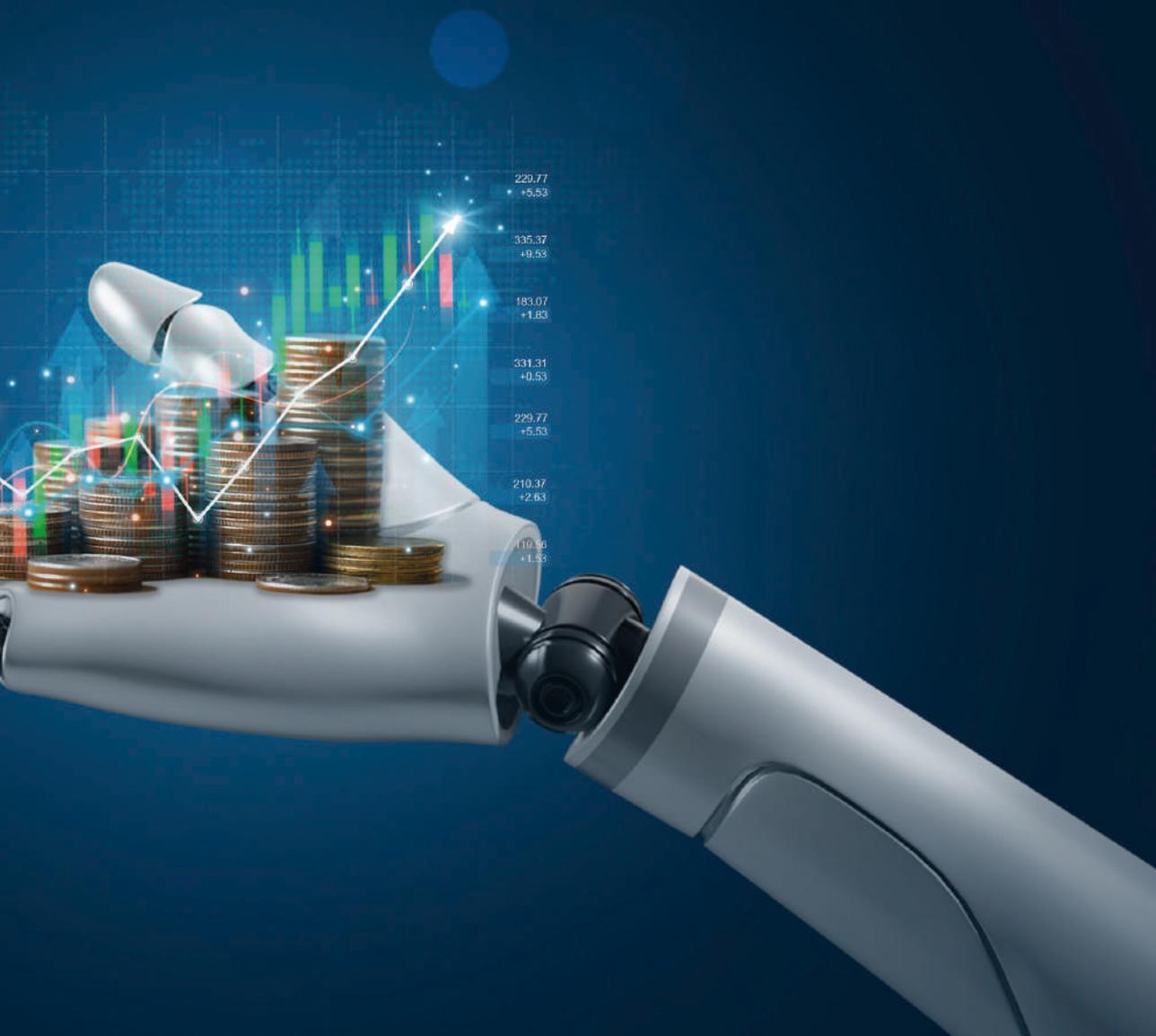
Balance de las elecciones de 2024

Las advertencias sobre la posible influencia de la IA generativa y, en particular, de los *deepfakes* en las elecciones de 2024 provienen tanto del caso del uso potencial como del rápido avance de la tecnología de IA en los últimos dos años, que ha hecho que la tecnología sea más accesible que en anteriores ciclos electorales importantes. Esas advertencias no han sido del todo exageradas.

Los *deepfakes* se han utilizado en las elecciones de 2024 para atacar a figuras políticas concretas, de varias formas diferentes. Se han utilizado audio y video generados por IA para crear grabaciones falsas de políticos. En las primarias demócratas de New



« La naturaleza auditiva o visual de algunos tipos de contenidos generados puede tener un mayor impacto emocional en las personas que la desinformación basada en texto. »



Hampshire, se utilizó audio generado por IA para llamadas automáticas con la voz de Biden instando a los votantes a no votar (Swenson y Weissert, 2024). Aun así, Biden ganó ampliamente las primarias de Nuevo Hampshire a pesar del intento de desalentar el voto.

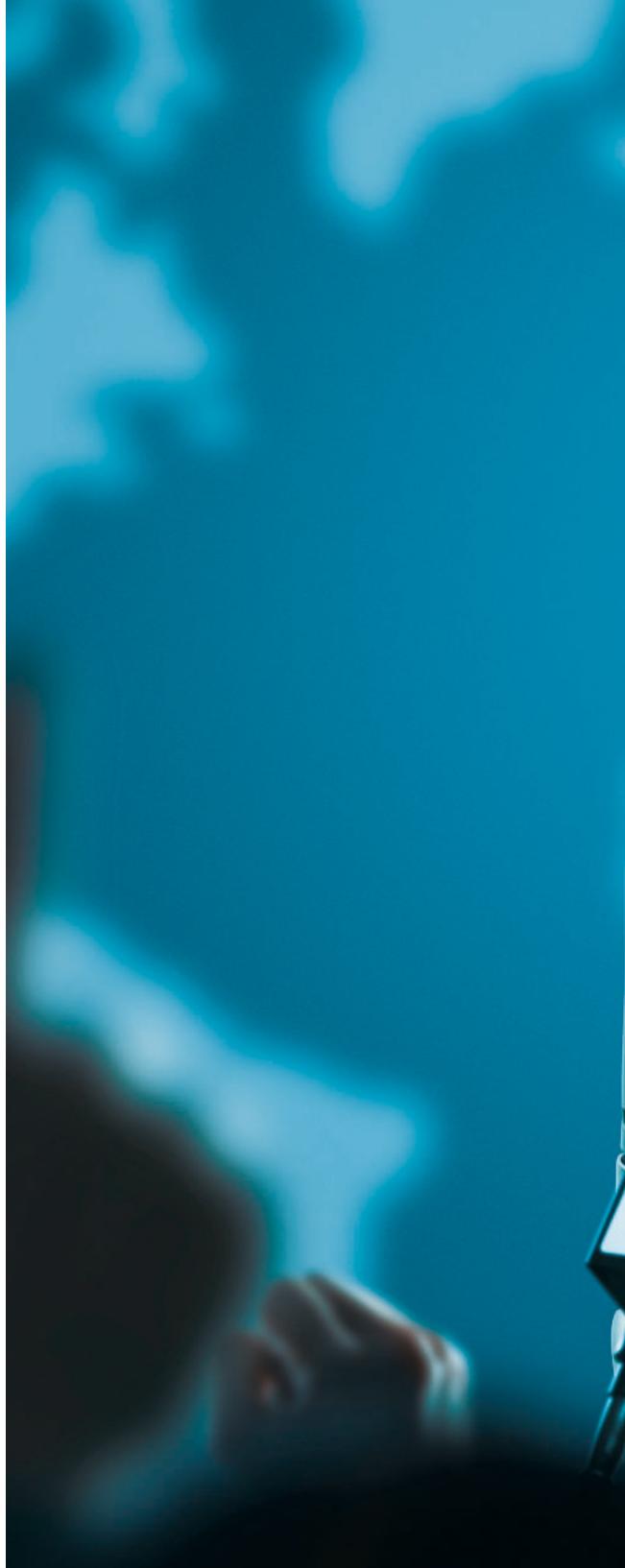
En Eslovaquia, un audio atribuyó falsamente planes para amañar unas elecciones a un líder político antes de las elecciones

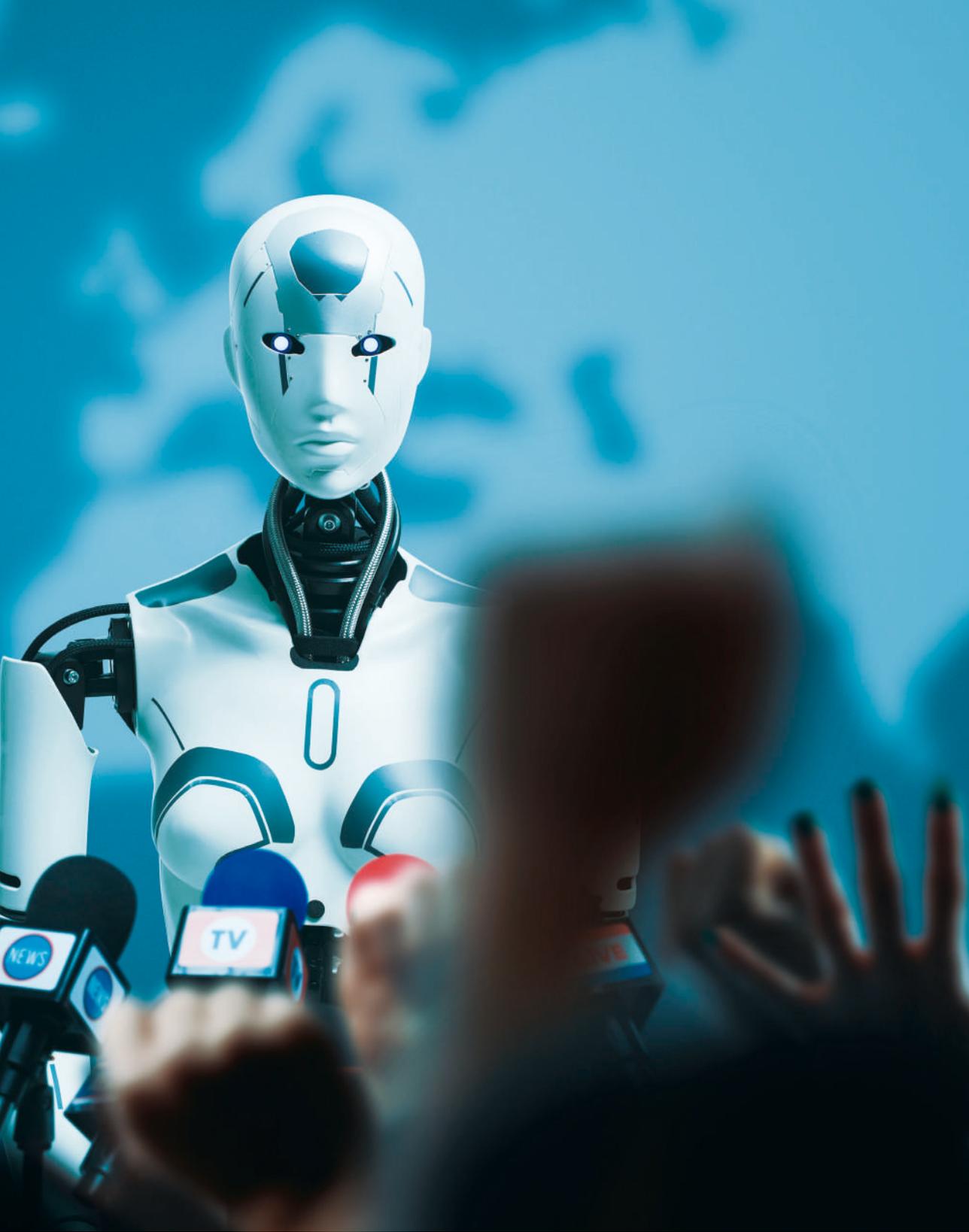
parlamentarias de 2023 (Devine, 2024). También se han utilizado imágenes y videos generados por IA para tergiversar la imagen de los políticos. Imágenes generadas por IA mostraban a Donald Trump con votantes negros con la intención de ganarse su simpatía (Sainato, 2024). Durante las primarias de 2023, un comité de acción política asociado con el gobernador de Florida, Ron DeSantis, utilizó audio generado por IA de

Donald Trump para mostrarlo atacando al gobernador de Iowa, con la intención de pintarlo como irrespetuoso con el estado de caucus (Isenstadt, 2023). Otros *deepfakes* se han utilizado para crear contenidos embarazosos o comprometedores. Por ejemplo, en el Reino Unido, una investigación descubrió 400 casos de *deepfakes* alterados digitalmente que mostraban a 30 políticos británicos de alto nivel en situaciones comprometidas (Morgan, 2024).

En Polonia, el partido de la oposición Plataforma Civil creó un video imitando la voz del primer ministro («Opposition criticized...», 2023). El contenido se basaba en correos electrónicos filtrados del jefe de gabinete del primer ministro y alternaba entre mostrar clips de video reales del primer ministro hablando y clips de audio generados por IA que leían secciones de los correos electrónicos filtrados. La motivación aparente era contrastar las declaraciones públicas del primer ministro sobre la unidad dentro de la coalición gobernante con mensajes privados que reconocían tensiones en el gobierno. Solo tras el escepticismo y las críticas, la plataforma reconoció que el contenido había sido generado por IA.

En el Reino Unido, el primer día de la conferencia del Partido Laborista en Liverpool se difundió un clip de audio que mostraba a Keir Starmer insultando y profiriendo blasfemias contra los miembros de su personal y otro que lo mostraba criticando a la ciudad de Liverpool. Los clips, publicados en X por una cuenta con menos de 3.000 seguidores, recibieron 1,4 millones de visitas («Deepfake audio...», 2023). Una empresa de detección de *deepfakes*, Reality Defender, indicó que era probable que el audio estuviera manipulado en un 75% y el análisis del gobierno británico también confirmó la falta de autenticidad del contenido (Martin, 2023). Aunque voces de todo





« Si la gente llega a pensar que cualquier cosa puede ser falsa, es posible que no confíe en que algo sea cierto. »

el espectro político criticaron el audio, este provocó preocupación por la amenaza que suponen los *deepfakes* para la democracia y puso de relieve los retos que plantean los contenidos que desacreditan.

La creación de *deepfakes* no se limita a la política nacional. Rusia ha sido acusada en repetidas ocasiones de crear *deepfakes* sobre la presidenta moldava prooccidental Maia Sandu para ridiculizar a la dirigente y socavar su credibilidad. CopyCop, una red presuntamente de influencia rusa, ha utilizado IA y sitios no auténticos para crear y diseminar desinformación. Sin embargo, los informes sugieren poca participación o amplificación en las redes sociales (Insikt Group, 2024).

Aunque estos ejemplos apuntan a formas en las que grupos o individuos han utilizado *deepfakes* generados por IA, las elecciones de 2024 también han destacado por la ausencia de *deepfakes*. En las elecciones de México de 2024, los *deepfakes* no aparecieron en absoluto y, de hecho, las elecciones parecen haberse desarrollado sin incidentes perturbadores importantes, aunque los actores externos tuvieron una causa poco probable (Pruneda y Salazar Ugarte, 2024).

A pesar de la disponibilidad de *deepfakes* y de las tensiones entre Taiwán y China, su uso fue limitado. China pareció utilizar clips de audio generados por IA para atacar al candidato presidencial del Partido Democrático Progresista, Lai Ching-te (Hung et al., 2024). El contenido generado por IA in-

cluía videos manipulados del candidato, en los que aparecía Lai hablando de escándalos que no se habían producido y apoyando una coalición con el Kuomintang (KMT), el partido nacionalista chino que gobernó hasta que fue derrotado por el Partido Comunista de China en el continente.

Pero más que *deepfakes*, China recurrió a otras técnicas de desinformación: difundió historias falsas o engañosas en las redes sociales, como la presentación de Estados Unidos como un aliado poco fiable que abandonaría a Taiwán, y presentó a los sufragios como una elección entre la paz (unificación con China) y la guerra (continuación de la independencia), difundiendo afirmaciones falsas sobre laboratorios biológicos estadounidenses en Taiwán, promoviendo teorías conspirativas sobre la injerencia de la CIA en las elecciones y promoviendo narrativas racistas contra los trabajadores migrantes.

China se ha asociado con técnicas como el *spamouflage* (camuflaje mediante *spam*), en el que grupos afiliados al gobierno chino utilizan cuentas para promover activamente determinadas narrativas en las redes sociales (Tsu, 2024). Por ejemplo, estas cuentas pretenden dar una imagen negativa de Estados Unidos destacando la decadencia urbana, la brutalidad policial y el deterioro de las infraestructuras. Estas cuentas son especialmente activas durante acontecimientos como catástrofes naturales o elecciones. En abril de 2023, el Departamento de Justicia de Estados Unidos acusó a 40 empleados del Grupo de Trabajo de Proyectos Especiales 912 del Ministerio de Seguridad Pública chino por su participación en una campaña de influencia que parecía ser *spamouflage* («40 Officers...», 2023).

Además de los continuos esfuerzos por camuflar el *spam*, China no solo distribuyó estos mensajes, sino que se apoyó en

representantes taiwaneses para difundir desinformación y dificultar su rastreo hasta China (Iyengar, 2024). Más allá de los resultados electorales inmediatos, China parecía centrada en erosionar con el paso del tiempo la confianza en la democracia y la soberanía de Taiwán.

Pero en ningún caso prevalecieron los *deepfakes* ni parecieron eficaces. Lo mismo puede afirmarse de las elecciones indias de 2024. A pesar de la preocupación por el uso generalizado de *deepfakes*, el número real de casos verificados de desinformación generada por IA fue relativamente bajo. De las 258 comprobaciones de hechos relacionados con las elecciones realizadas por Boom Live, solo en doce hubo desinformación generada por inteligencia artificial (Jacob, 2024). Los dos principales partidos, el BJP y el Partido del Congreso, crearon y compartieron contenidos generados por inteligencia artificial, como memes y videos satíricos, discursos traducidos por inteligencia artificial o llamadas robotizadas personalizadas. Un estudio demostró que, de unos dos mil mensajes virales de WhatsApp, solo el 1% fueron generados por IA, una pequeña huella según la revista *Nature* (Garimella y Chauchard, 2024). La prevalencia y el impacto del mal uso de los contenidos generados por IA fueron limitados, sin duda, menos de lo que se temía en un principio.

¿Por qué los contenidos generados por IA han tenido tan poco impacto?

Los estudiosos han mostrado pruebas de concepto de cómo los actores maliciosos podrían utilizar la desinformación generada por IA a gran escala para perturbar las elecciones democráticas. Sin embargo, escasean las pruebas que sugieran que los actores están produciendo dicho contenido o tienien-

do impacto en las elecciones, lo que plantea la cuestión de por qué las predicciones han estado en desacuerdo con la realidad (Pawelec, 2022).

Una de las razones de su escasa eficacia es que la tecnología aún está en sus inicios, sobre todo en lo que respecta a los videos o audios. Los usuarios identifican y desacreditan rápidamente las imágenes o el video, en parte, porque los *deepfakes* aún son identificables. La incorporación de video y audio a veces crea una sincronización desajustada en la forma en que se proyectan las palabras, lo que significa que la atención a los movimientos de los labios puede poner de manifiesto incoherencias. Por otra parte, a menudo hay incoherencias sutiles entre la IA y la imagen humana, como ocurrió con las falsificaciones de Tom Cruise (Fisher, 2022). Las personas que ven estas imágenes pueden experimentar una respuesta emocional escéptica conocida como el valle inquietante al ver una imagen robótica que sutilmente no se parece a la realidad (Gillis, 2024).

Otro conjunto de factores es que, aunque la tecnología es incipiente, los individuos, los líderes políticos, los Estados y las plataformas de medios sociales están realmente preparados para estos *deepfakes*, lo que ayuda a neutralizar el efecto (Ellison y Wingett Sanchez, 2024). En algunos casos, los funcionarios públicos han realizado simulacros y ejercicios de simulación para responder a los *deepfakes*. Algunos Estados han aprobado leyes que regulan el uso de *deepfakes* en campañas políticas, lo que puede disuadir a algunos posibles malos actores (Polgár y Wen, 2024). La detección algorítmica está mejorando de forma que permite a las plataformas de las redes sociales identificar y eliminar los *deepfakes*, muchos de los cuales han sido prohibidos por estas plataformas (De Vynck, 2024).

Otra consideración, como sugiere el ejemplo de China en las elecciones taiwa-

nesas de 2024, es que los actores pueden simplemente ver más prácticas o eficaces otras formas de desinformación, tales como aparecer en foros en línea con perspectivas particulares o con valencias políticas para moldear lo que la gente lee.

Además, la persuasión política es difícil. La investigación sugiere que el efecto de la misinformación o la desinformación (con intención de engañar) tiende a tener repercusiones limitadas porque las opiniones de la gente están bastante arraigadas. De hecho, estudios sobre si la desinformación ha afectado al comportamiento político o cómo lo ha hecho suelen arrojar resultados nulos porque la gente tiende a no cambiar de opinión, incluso cuando se enfrenta a las imágenes más visceralmente poderosas (frente al texto) (Adam, 2024).

Conclusión

A pesar de su potencial perturbador, los *deepfakes* y los contenidos generados por IA en general aún no son tan frecuentes ni han tenido tanta repercusión durante las oleadas electorales de 2024. Los esfuerzos tanto para identificar como para eliminar las falsificaciones han aumentado en particulares, gobiernos y plataformas. El impacto en la opinión pública ha sido limitado. Las medidas que la sociedad ha adoptado para protegerse de los *deepfakes* han mitigado su impacto, pero el éxito hasta ahora en la mitigación de la influencia de los *deepfakes* no garantiza la inmunidad futura.

No obstante, a medida que evolucione la IA también lo harán los esfuerzos por manipular los sistemas democráticos para obtener ventajas electorales. La conciencia pública y la alfabetización mediática seguirán desempeñando un papel importan-

te en la reducción del impacto. A medida que cambie la tecnología, también deberán hacerlo las campañas de los gobiernos, las organizaciones sin ánimo de lucro y los medios de comunicación para poner de relieve el tipo de pensamiento crítico y el escepticismo necesarios para consumir contenidos digitales con el fin de protegerse contra la manipulación y la desinformación. Los medios de comunicación tendrán que seguir verificando rigurosamente la autenticidad de los contenidos visuales o sonoros antes de su publicación, para evitar la propagación de *deepfakes* y reforzar la confianza en los medios. Los marcos normativos y jurídicos tendrán que mantenerse al día y seguir evaluando las nuevas tecnologías y preguntándose qué tipos de *deepfakes* son libertad de expresión y cuáles deben prohibirse.

De este análisis se desprende una advertencia adicional. Como sugiere la influencia *online* china en las elecciones taiwanesas, los actores que buscan influir en la opinión pública seguirán utilizando otros medios, como ejércitos de *trolls online* o grupos que se dedican al *spamouflage*, creando mensajes destinados a manipular la opinión política. La preocupación por los *deepfakes* puede oscurecer esos enfoques y, lo que es peor, desviar los recursos y la atención no solo de esos mecanismos más establecidos que incluyen el *phishing*, el *ransomware* y otras ciberamenazas. Las medidas de seguridad, la financiación y los esfuerzos de investigación podrían inclinarse desproporcionadamente hacia la lucha contra los *deepfakes* a expensas de iniciativas de ciberseguridad más amplias, un costo de oportunidad en términos de tiempo y esfuerzo invertidos en la detección y prevención de *deepfakes*. Sin embargo, un énfasis excesivo en los *deepfakes* puede contribuir a un proceso auto-

» Las democracias deben fomentar una cultura de pensamiento crítico y alfabetización mediática entre sus ciudadanos. «

cumplido, en forma de erosión de la confianza en los medios de comunicación y la información, al inculcar escepticismo en el público y poner en duda los contenidos legítimos y auténticos.

Los continuos avances en las tecnologías de detección, la mayor concientización pública y los sólidos marcos jurídicos han demostrado su eficacia a la hora de mitigar muchas de las amenazas que plantean los *deepfakes*. Sin embargo, la naturaleza dinámica de la tecnología y la constante evolución de las tácticas de los actores maliciosos requieren una vigilancia y una adaptación continuas. Las democracias deben fomentar una cultura de pensamiento crítico y alfabetización mediática entre sus ciudadanos, manteniendo al mismo tiempo la transparencia y la responsabilidad de sus instituciones. De este modo, pueden salvaguardar su integridad y seguir prosperando en la era digital, convirtiendo las posibles vulnerabilidades en fortalezas y garantizando que sus valores fundacionales perduren frente a los retos de la tecnología moderna. ♦

Referencias bibliográficas

40 Officers of China's National Police Charged in Transnational Repression Schemes Targeting U.S. Residents. (2023, 17 de abril). *Office of Public Affairs*. <https://www.justice.gov/opa/pr/40-officers-china-s-national-police-charged-transnational-repression-schemes-targeting-us>

ADAM, D. (2024, 18 de junio). Misinformation might sway elections — but not in the way that you think. *Nature*. <https://www.nature.com/articles/d41586-024-01696-z>

BICKERTON, J. (2023, 24 de marzo). Deepfakes Could Destroy the 2024 Election. *Newsweek*. <https://www.newsweek.com/deepfakes-could-destroy-2024-election-1790037>

CAHLAN, S. (2020, 13 de febrero). How misinformation helped spark an attempted coup in Gabon. *The Washington Post*. <http://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/>

DE VYNCK, G. (2024, 5 de abril). The AI apocalypse is here. These are the ideas for fighting it. *The Washington Post*. <https://www.washingtonpost.com/technology/2024/04/05/ai-deepfakes-detection/>

Deepfake audio of Sir Keir Starmer released on first day of Labour conference. (2023, 9 de octubre). *Sky News*. <https://news.sky.com/story/labour-faces-political-attack-after-audio-is-posted-of-sir-keir-starmer-12980181>

DEVINE, C., O'SULLIVAN, D., Y LYNGAAS, S. (2024, 1 de febrero). A fake recording of a candidate saying he'd rigged the election went viral. Experts say it's only the beginning. *CNN*. <https://edition.cnn.com/2024/02/01/politics/election-deepfake-threats-invs/index.html>

ELLIOT, V. (2024, 30 de mayo). 2024 Is the Year of the Generative AI Election. *Wired*. <https://www.wired.com/story/2024-is-the-year-of-generative-ai-elections/>

ELLISON, S., Y WINGETT SANCHEZ, Y. (2024, 8 de mayo). In Arizona, election workers trained with *deepfakes* to prepare for 2024. *The Washington Post*. <https://www.washingtonpost.com/politics/2024/05/08/arizona-election-workers-trained-with-deepfakes-prepare-2024/>

- EWE, K. (2023, 28 de diciembre). The Ultimate Election Year: All the Elections Around the World in 2024. *Time*. <https://time.com/6550920/world-elections-2024/>
- FISHER, M. (2022, 21 de julio). How I Became the Fake Tom Cruise. *Hollywood Reporter*. <https://www.hollywoodreporter.com/feature/deepfake-tom-cruise-miles-fisher-1235182932/>
- GARIMELLA, K. Y CHAUCHARD, S. (2024, 5 de junio). How prevalent is AI misinformation? What our studies in India show so far. *Nature*. <https://www.nature.com/articles/d41586-024-01588-2>
- GILLIS, A. (2024, febrero). uncanny valley. *Techtarget*. <https://www.techtarget.com/whatis/definition/uncanny-valley>
- GOLDSTEIN, J., CHAO, J., GROSSMAN, S., STAMOS, A., Y TOMZ, M. (2024, 20 de febrero). How persuasive is AI-generated propaganda? *PNAS Nexus*. <https://doi.org/10.1093/pnasnexus/pgae034>
- Here's How Deepfakes, Like the One BJP Used, Twist the Truth. (2020, 20 de febrero). *VICE*. <https://www.vice.com/en/article/bjp-manoj-tiwari-deepfake-twists-truth/>
- HUNG, CH-L., FU, W.-CH., LIU, CH-C., Y TSA, H-J. (2024). AI Disinformation Attacks and Taiwan's Responses during the 2024 Presidential Election. *Thomson Foundation*. https://www.thomsonfoundation.org/media/268943/AI_disinformation_attacks_taiwan.pdf
- INSIKT GROUP. (2024, 24 de junio). Russia-Linked CopyCop Expands to Cover US Elections, Target Political Leaders. *Recorded Future*. <https://www.recordedfuture.com/research/copycop-expands-to-cover-us-elections-target-political-leaders>
- ISENSTADT, A. (2023, 17 de julio). DeSantis PAC uses AI-generated Trump voice in ad attacking ex-president. *Politico*. <https://www.politico.com/news/2023/07/17/desantis-pac-ai-generated-trump-in-ad-00106695>
- IYENGAR, R. (2024, 23 de enero). How China Exploited Taiwan's Election—and What It Could Do Next. *FP*. <https://foreignpolicy.com/2024/01/23/taiwan-election-china-disinformation-influence-interference/>
- JACKSON, K., SCHIFF, D, Y BUENO, N. (2024, 20 de febrero). The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability? *American Political Science Review*, First View, pp. 1-20. <https://doi.org/10.1017/S0003055423001454>
- JACOB, N. (2024, 3 de junio). 2024 Elections Report: Fake Polls, Cheap Voice Clones, Communal Claims Go Viral. *Boom*. <https://www.boomlive.in/boom-research/2024-elections-report-fake-polls-cheap-voice-clones-communal-claims-go-viral-25532>
- KREPS, S. et al. (2020, noviembre). All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science*, 9(1), 1-14. <https://doi.org/10.1017/XPS.2020.37>
- KREPS, S., Y KRINER, D. (2023, octubre). How AI Threatens Democracy. *Journal of Democracy*. www.journalofdemocracy.org/articles/how-ai-threatens-democracy/
- MARTIN, A. (2023, 9 de octubre). UK opposition leader targeted by AI-generated fake audio smear. *The Record*. <https://therecord.media/keir-starmer-labour-party-leader-audio-smear-social-media-deepfake>
- MARTINAU, K. (2023, 20 de abril). What is generative AI? *IBM*. <https://research.ibm.com/blog/what-is-generative-ai>
- MORGAN, L. (2024, 2 de julio). Deepfake pornography is being used to humiliate and

silence powerful female politicians like Angela Rayner and Penny Mordaunt. Why doesn't the law protect them? *Glamour*. <https://www.glamourmagazine.co.uk/article/female-mps-politicians-deepfake-victims>

Opposition criticised for using AI-generated deepfake voice of PM in Polish election ad (2023, 25 de agosto). *Notes from Poland*. <https://notesfrompoland.com/2023/08/25/opposition-criticised-for-using-ai-generated-deepfake-voice-of-pm-in-polish-election-ad/>

PAWELEC, M. (2022, septiembre). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*, 1(2).

POLGÁR, J., Y WEN, A. (2024, 10 de octubre). Deceptive Audio or Visual Media ('Deepfakes') 2024 Legislation. *NCSL*. <https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation#:~:text=Beginning%20in%202019%2C%20several%20states>

PRUNEDA, P., Y SALAZAR UGARTE, P. (2024, 29 de mayo). Elections in Mexico: Beyond "Deepfakes". *Wilson Center*. <https://www.wilsoncenter.org/article/elections-mexico-beyond-deepfakes>

SAINATO, M. (2024, 4 de marzo). AI-generated images of Trump with Black voters being spread by supporters. *The Guardian*. <https://www.theguardian.com/us-news/2024/mar/04/trump-ai-generated-images-black-voters>

SWENSON, A., Y WEISSERT, W. (2024, 23 de enero). New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary. *Ap*. <https://apnews.com/article/new-hampshire-primary-biden-ai-deepfake-roboca>

ll-f3469ceb6dd613079092287994663db5
TSU, T. (2024, 15 de febrero). Chinese Influence Campaign Pushes Disunity Before U.S. Election, Study Says. *The New York Times*. <https://www.nytimes.com/2024/02/15/business/media/chinese-influence-campaign-division-elections.html>

VERMA, P. (2023, 17 de diciembre). The rise of AI fake news is creating a "misinformation superspreader". *The Washington Post*. <https://www.washingtonpost.com/technology/2023/12/17/ai-fake-news-misinformation/>

YASAR, K. (2024). What is deepfake technology? *Techtarget*. <https://www.techtarget.com/whatis/definition/deepfake>

Traducción inglés-español: Manfred Steffen.



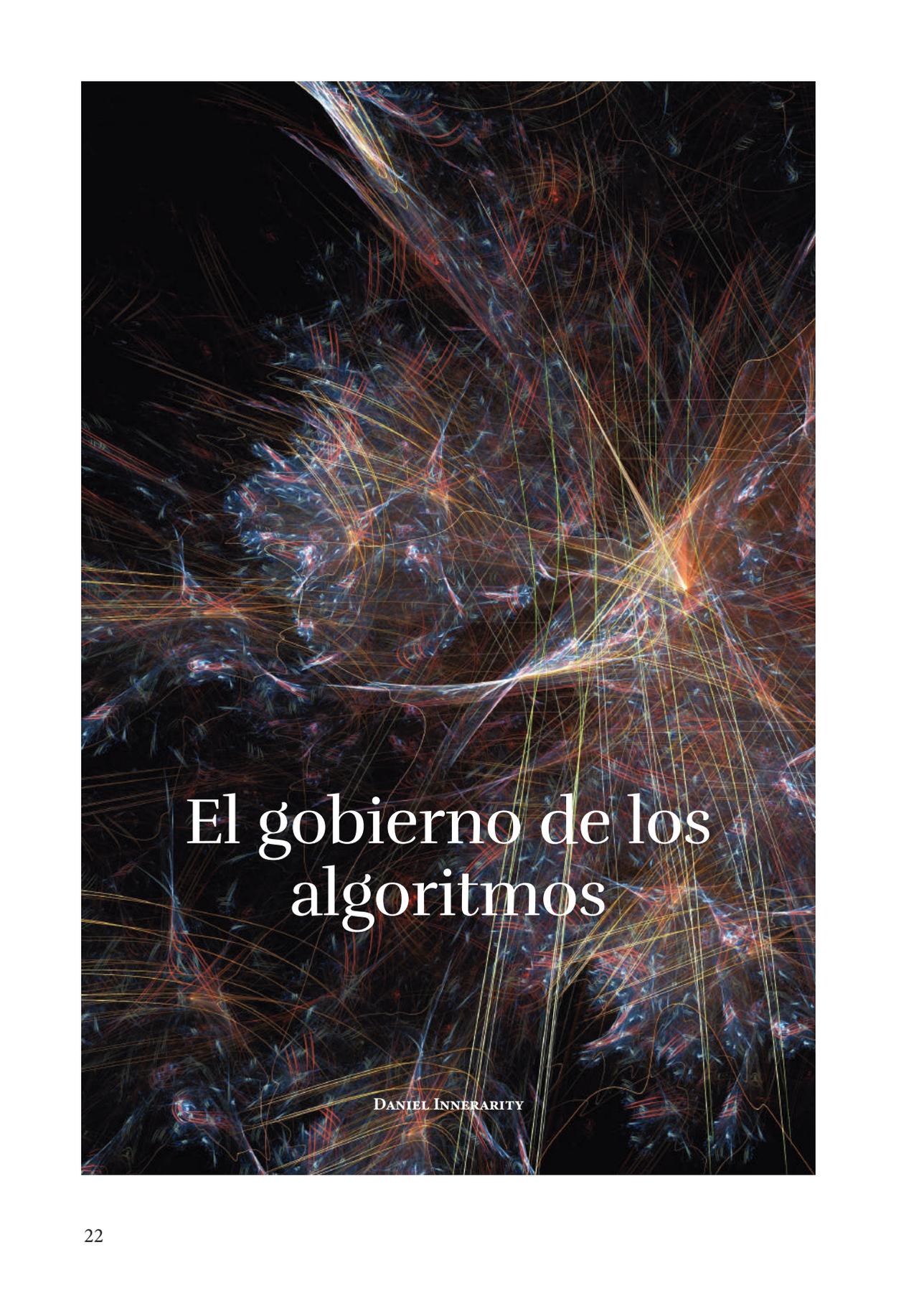
Sarah E. Kreps

Politóloga estadounidense. Veterana de la fuerza aérea. Analista de políticas sobre política exterior y de defensa de los Estados

Unidos. Profesora de gobierno en la Universidad de Cornell y profesora adjunta en el Instituto de Guerra Moderna de West Point. Autora de libros sobre drones, intervenciones militares de los Estados Unidos y cómo Estados Unidos financia sus guerras. Miembro vitalicio del Consejo de Relaciones Exteriores.

X: @sekreps

LI: sarah-kreps-51a3b7257



El gobierno de los algoritmos

DANIEL INNERARITY

Una cuestión es cómo gobernar los algoritmos y otra bien distinta es si los algoritmos llegarán un día a goberarnos. Tenemos que examinar las expectativas de la gobernanza algorítmica y sus límites. No parece que los algoritmos puedan hacerse cargo de todo el proceso político, con la eficacia que prometen y con la legitimidad que debería justificar ese nuevo régimen.

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.

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 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 sea personal, indiosincrático o perspectivista; ya no se confía en la in-

tegridad de los que dicen la verdad o en el prestigio de instituciones ejemplares sino en procedimientos fuertemente estandarizados (Porter, 1995). 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, p. 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 estas.

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, p. 71); «la autoridad se expresa cada vez más algorítmicamente» (Pasquale, 2015, p. 1).

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 apoyan a los humanos en sus decisiones e incluso los sustituyen, 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) es compatible con lo que consideramos un sistema político de toma de decisiones. 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.

Expectativas 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, pero que también tienen sus límites e inconvenientes.

a. La promesa de objetividad

Resulta muy seductora la promesa de la decisión algorítmica: no se trata de ahorrar tiempo y dinero, sino de promover la objetividad. 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, p. 79). 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. 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.

Esta pretensión no es del todo nueva, al igual que tampoco lo es su crítica. La idea de autoridad burocrática de Weber ya había ensalzado los valores de eficiencia y objetividad, pero 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. 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 esa promesa, que generan otro tipo de distorsiones, que no están exentas de arbitrariedad y que los algoritmos a menudo reflejan e incluso potencian los prejuicios que están profundamente asentados en una sociedad.

b. 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; se reforzaría así 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. Con un mundo lleno de sensores, algoritmos, datos y objetos inteligentes se configura una suerte de *sensorium* social que permite personalizar la salud, los transportes o la 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. A la objetividad de los métodos de gobernanza algorítmica le correspondería una mayor subjetividad en sus destinatarios, que verían así mejor conocida, respetada y satisfecha su particularidad.

El cómodo paternalismo de las sociedades algorítmicas consiste en que da a las personas lo que estas quieren, que gobierna con incentivos proporcionados, que 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 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. »

la satisfacción de necesidades se hace con frecuencia a cambio de espacios de libertad. Es cierto que así se satisfacen muchos de nuestros deseos, pero 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.

Limitaciones de la gobernanza algorítmica

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.

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

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. 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.



» Cuestiones políticas son aquellas que solo se pueden resolver con juicios de valor. «

En este y otros casos, no se trata solo de implementación o transparencia de los algoritmos utilizados, sino de juicios de valor en torno a las posibilidades alternativas de definir los objetivos de la educación, que son diversos y en parte concurrentes, como corresponde a una sociedad pluralista. Los procesos de negociación política tienen



« La decisión acerca de qué es computarizable no se puede a su vez computarizar. »

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 aquellas cuestiones técnicas en las que se decide la imple-

mentación técnica de los objetivos pretendidos y sobre la base del saber disponible. En ocasiones también es algo políticamente controvertido qué clase de optimización es satisfactoria y qué 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.

La gobernanza algorítmica se orienta a realizar objetivos que no han sido discutidos, 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. 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 ha de servir. Podría formularse esta idea recordando aquella afirmación de John von Neumann: 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 (Neumann, 1966, p. 51). Dicho de otra manera: la decisión acerca de qué es computarizable no se puede a su vez computarizar.

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. Los artefactos que se encargan de optimizar las decisiones no hacen innecesaria la discusión acerca de qué es una buena decisión. Es cierto que la inteligencia artificial sirve para informar decisiones y optimizar resultados,

» La gran promesa de la gobernanza algorítmica es que unos resultados óptimos nos hagan olvidar los procedimientos deseables. «

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 política.

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.

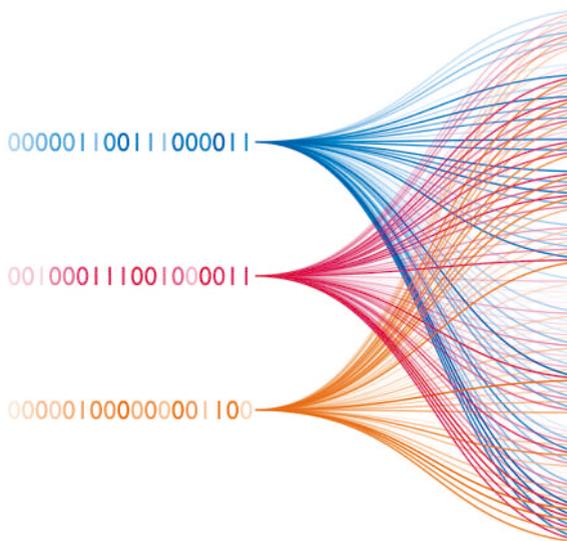
La inevitabilidad de decidir

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 dejar de insistir en unas prácticas que no se corresponden en absoluto con las nuevas realidades sociales. Estamos ante un in-

tento de concebir la sociedad de un modo despolitizado.

Las sociedades contemporáneas necesitan una enorme movilización cognitiva para hacer frente a los problemas que deben resolver, pero el argumento último a favor de la democracia no es epistémico sino decisional. 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 inteligencia artificial no pueden exonerarnos de esa





« La legitimidad final [de las sociedades] 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. »

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. Un mundo humano tiene que ser un mundo negociable. ♦

Referencias

DASTON, L., Y GALISON, P. (2010). *Objectivity*. Princeton University Press.

GILLESPIE, T. (2014). The Relevance of Algorithms. En T. GILLESPIE, P. J. BOCZKOWSKI y K. A. FOOT (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167-193) Cambridge: The MIT Press.

KRÜGER, J., Y LISCHKA, K. (2018). Was zu tun ist, damit Maschinen den Menschen dienen. En R. MOHABBAT KAR, B. THAPA y P. PARYCEK (eds.), *(Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft* (pp. 440-470). Berlín: Fraunhofer-Institut für Offene Kommunikationssysteme fokus.

LASH, S. (2007). Power after hegemony. *Theory, Culture & Society*, 24(3), 55-78.

NEUMANN, J. VON (1966). *Theory of Self-Reproducing Automata*. Urbana: University of Illinois Press.

PASQUALE, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press.

PORTER, T. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.

SUCHMAN, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions*, Cambridge University Press.



Daniel Innerarity

Catedrático de filosofía política. Investigador Ikerbasque de la Fundación Vasca para la Ciencia en la Universidad del País

Vasco, España. Titular de la cátedra de inteligencia artificial y democracia, en el Instituto Europeo de Florencia, Italia.

X: @daniinnerarity

LI: daniel-innerarity-406665112



Gobernanza anticipatoria con ayuda de la tecnología

LYDIA GARRIDO LUZARDO

El poder transformador y el potencial disruptivo de la IA requieren de una gobernanza ética. La *anticipación responsable* y el uso alfabetizado del futuro son elementos clave en la formulación de políticas. Las comisiones parlamentarias del futuro ofrecen recomendaciones para una gobernanza inclusiva y basada en la cooperación, la transparencia y la inteligencia colectiva.

El desarrollo de la inteligencia artificial ha alcanzado un punto de inflexión en el que su capacidad transformativa y disruptiva exige una reflexión profunda sobre cómo gobernarla de manera ética y responsable. El concepto de inteligencia artificial general (AGI, por sus siglas en inglés) refiere a un tipo de IA capaz de realizar cualquier tarea intelectual humana y plantea desafíos sin precedentes que requieren otros marcos epistémicos y metodológicos para su gobernanza. Estos marcos deben ser capaces de no solo reaccionar a los cambios tecnológicos luego de ocurridos, sino también anticipar los riesgos y oportunidades asociados con la evolución de la IA, dentro de una concepción más amplia, de complejidad sistémica anticipatoria apoyada en inteligencia colectiva y valores éticos.

En este artículo se propone que una gobernanza anticipatoria efectiva requiere un enfoque fundamentado en capacidades y competencias anticipatorias, así como en principios éticos sólidos. La *anticipación responsable* se presenta aquí como una capacidad y cualidad clave para la toma de decisiones previsoras y proactiva, creando las condiciones necesarias para que el desarrollo de la IA evolucione de manera ética, controlada y beneficiosa para toda la sociedad. Esto implica establecer principios que orienten su

evolución hacia una AGI segura y ética, y un enfoque que permita gestionar sus efectos futuros *ex ante*.

El texto articula aspectos de la ética, la complejidad y el *uso del futuro* en la toma de decisiones, aplicándolos al contexto de la gobernanza anticipatoria de la IA. También se incluye la experiencia de la Comisión Especial de Futuros del Parlamento del Uruguay y los aportes recientes de la Segunda Cumbre Mundial de Comisiones de Futuros en Parlamentos.

El trabajo se estructura en tres secciones principales. En primer lugar, se analiza la naturaleza evolutiva de la IA y su potencial disruptivo. En segundo lugar, se exploran los desafíos que el *uso del futuro* plantea para una práctica de gobernanza anticipatoria. Finalmente, se discuten las consideraciones prácticas para una gobernanza responsable de la IA, destacando el papel crucial que desempeñan los Parlamentos y otras instituciones en el diseño de marcos flexibles, anticipatorios y adaptativos.

La naturaleza evolutiva de la IA y su alcance disruptivo

Hay múltiples definiciones para lo que se ha dado en llamar *inteligencia artificial*.

Algunos debates se centran en las diferencias con la inteligencia humana. En este texto no nos enfocaremos en este aspecto. Su potencial disruptivo va más allá de las similitudes o diferencias (que generalmente son discutidas desde una visión antropomórfica y antropocéntrica); en cambio, colocamos la cuestión en su alcance, es decir, en sus actuales y potenciales capacidades que pueden ser oportunidades para el bien de la humanidad o graves amenazas.

El otro aspecto que interesa es el carácter *evolutivo* de la IA (y esto es justamente lo que las más de las veces genera esas diferencias y falta de consenso sobre una única definición), ya que una característica inherente es su cambio permanente. La OCDE modificó en diciembre de 2023 su definición para los sistemas de inteligencia artificial:

Un sistema de IA es un sistema basado en una máquina que para un conjunto de objetivos explícitos o implícitos infiere a partir de la entrada que recibe cómo generar salidas en la forma de predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los diferentes sistemas de IA varían en sus niveles de autonomía y adaptabilidad después de la implementación¹.

De manera simplificada, podemos decir que la IA no es estática; está en constante evolución, expandiendo sus capacidades y modificando su relación con las dinámicas tecnológicas y sociales que la rodean. Este proceso evolutivo, caracterizado por una autonomía creciente y una capacidad disruptiva, plantea desafíos complejos para los tomadores de de-

¹ Uno de los ajustes principales que se hicieron a esta definición fue quitarle que tanto las entradas como las reglas pueden ser colocadas o no por un humano; en cambio, se reconoce que el sistema es capaz de aprender nuevos objetivos por sí mismo.

cisión, y cuestiona las capacidades tradicionales de los sistemas de gobernanza. Además, este proceso se entrelaza con otras tecnologías emergentes como el internet de las cosas (IoT), los sistemas autónomos, la robótica, la bio y nanotecnología, y las ciencias cognitivas y neurociencias, lo que amplifica aún más su impacto. Un ejemplo de este entrelazamiento se observa en la medicina de precisión, donde la IA, junto con la biotecnología y la nanotecnología, permite desarrollar tratamientos personalizados, pero también plantea desafíos éticos y de privacidad que requieren nuevas formas de gobernanza.

Desde sus inicios, la IA ha demostrado ser una herramienta poderosa para resolver problemas específicos. Sin embargo, estamos a punto de un salto hacia la inteligencia artificial general, un tipo de IA con la capacidad de realizar tareas generales de manera similar a los humanos, incluyendo la resolución de problemas complejos sin preprogramación específica. Esta transición plantea preguntas fundamentales sobre cómo debemos gobernar estas tecnologías y qué tipo de futuros deseamos construir en relación con ellas.

Comprender la naturaleza evolutiva de la IA es crucial para diseñar marcos de gobernanza que permitan gestionar su impacto. Se tomará como base la distinción que hace el Millennium Project, Futures Studies and Research² de distintos tipos de IA.

IA estrecha, IA general y superinteligencia artificial

La gobernanza de la IA debe partir de una distinción fundamental entre tres tipos de IA: la inteligencia artificial estrecha (*Artificial*

² *Think Tank* con más de setenta nodos país de profesionales futuristas que se originó como grupo asesor para lo que fueron los primeros Objetivos del Milenio. Actualmente lleva adelante un estudio Delphi sobre *Global Artificial General Intelligence Governance*. <https://www.millennium-project.org/seeking-delphi-on-ai-implications-and-regulation-april-9-2024/>

Narrow Intelligence, ANI), la inteligencia artificial general (artificial general intelligence, AGI) y la superinteligencia artificial (*Artificial Super Intelligence*, ASI). Esta distinción es crítica porque cada una de estas formas de IA presenta riesgos, oportunidades y desafíos diferentes.

- ANI (*inteligencia artificial estrecha*) es el tipo de IA que tenemos en la actualidad, diseñada para realizar tareas específicas como el reconocimiento facial, la conducción autónoma o los sistemas de recomendación en plataformas de streaming. Estas ANI son herramientas limitadas en su capacidad y no pueden operar fuera de los parámetros para los que fueron diseñadas. La inteligencia artificial generativa (como el ChatGPT o similares) está demostrando en sus avances actuales estar en transición hacia un tipo o fase siguiente.
- AGI (*inteligencia artificial general*) representa un salto cualitativo en la evolución de la IA. A diferencia de la ANI, la AGI tiene la capacidad de resolver problemas no específicos, operar en una amplia gama de contextos sin intervención humana constante, adaptarse a nuevas situaciones y aprender de manera autónoma. Esto convierte a la AGI en un agente autónomo que podría actuar de maneras no anticipadas por los humanos que la desplieguen. La AGI puede mejorar su propio código y evolucionar rápidamente, lo que plantea preocupaciones sobre su control y las consecuencias imprevistas de su desarrollo. Los desarrollos actuales sugieren que la AGI podría estar próxima (los expertos discuten entre cinco años a un par de décadas para que se llegue a esa fase), y su impacto será profundo, ya que permitirá que los sistemas de IA actúen como agentes autó-

» Comprender la naturaleza evolutiva de la IA es crucial para diseñar marcos de gobernanza que permitan gestionar su impacto. «

nomos con capacidades comparables, o incluso superiores, a las humanas.

- ASI (*superinteligencia artificial*) es la evolución posterior a la AGI, una inteligencia tan avanzada que podría establecer sus propios objetivos y actuar de manera completamente independiente de los humanos. Aunque la ASI aún es una especulación, la posibilidad de que emerja de la AGI exige una gobernanza anticipatoria que contemple no solo las fases actuales del desarrollo de la IA, sino también sus posibles derivas.

En pocas palabras, mientras que la IA actual presenta retos en la transparencia y equidad de los algoritmos y de la privacidad personales, la AGI y la ASI amplifican estos desafíos, introduciendo cuestiones existenciales sobre el control, la autonomía y la influencia potencial en la estructura de la sociedad y humanidad (Glenn y Garrido, 2023).

Complejidad relacional y la necesidad de un cambio paradigmático

El carácter evolutivo de la IA, particularmente en su transición hacia la AGI y la ASI, requiere un cambio paradigmático en la forma de abordar su gobernanza. Es necesario pasar de un enfoque lineal y simplista a uno basado en la complejidad relacional, donde se reconozcan las múltiples interdependencias entre la tecnología y el

resto de diversos sistemas sociales y naturales (económicos, culturales, medioambientales, entre otros). La gobernanza no puede limitarse a la regulación reactiva de los problemas actuales o a intervenir después de que ocurran; debe ser anticipatoria y abrir alternativas en el presente mediante acciones preventivas y generativas, creando condiciones consensuadas socialmente que consideren el cambio cualitativo de estas tecnologías en su evolución. Esto requiere modelos participativos inter y transdisciplinarios, como modos reticulares (en red) basados en la inteligencia colectiva y la participación social.

« La gobernanza debe ser anticipatoria y abrir alternativas en el presente mediante acciones preventivas y generativas. »

Este enfoque de gobernanza anticipatoria no solo aborda los riesgos, sino que también explora las oportunidades que presenta la IA, y permite a los tomadores de decisiones gestionar crisis y, más importante aún, tener incidencia para transformar antes (ya sea evitando o creando condiciones deseadas).

La creación de la Comisión Especial de Futuros en el Parlamento de Uruguay es un claro ejemplo de este cambio. Al adoptar un enfoque más complejo e interrelacionado, la comisión trabaja para integrar perspectivas diversas e introducir aspectos del futuro, propulsando así acciones para un innovativo ecosistema de gobernanza anticipatoria.

Desafíos de la gobernanza anticipatoria y la anticipación responsable

Se propone una reflexión para la gobernanza anticipatoria de la IA desde la práctica de una *anticipación responsable*. El futuro en sí mismo no existe, lo cual coloca una dificultad ontológica que debe ser tomada en cuenta dentro de la dimensión epistémica del problema. Dado el carácter evolutivo de la IA, es vital no solo enfocarse en los problemas actuales, sino en incluir todas las fases posibles de desarrollo de la IA dentro de las estrategias de gobernanza. Sin embargo, surge una pregunta fundamental: ¿cómo asumir responsabilidad por futuros que aún no existen? A continuación, se plantea una reflexión y un abordaje para esta cuestión.

Ética y el imperativo de responsabilidad

La *anticipación responsable* busca articular una postura ética que vaya más allá de los enfoques puramente filosóficos o aquellos limitados a la previsión o prospectiva tradicional. En este contexto, se ubica la práctica ética en una conjunción consciente y reflexiva que permite *traer el futuro al presente* en diversas fases del proceso de toma de decisiones. Esto implica abrirlo para incluir el futuro desde la formulación de problemas, la construcción de alternativas y la selección de acciones, todo ello basado en insumos que estén alineados en *clave de futuros* (Garrido, 2024).

El trabajo pionero de Hans Jonas en la década de 1970 subrayó la importancia de integrar las implicaciones futuras en las decisiones éticas contemporáneas. Jonas desafiaba los estándares éticos predominantes al introducir un marco orientado al futuro que consideraba el impacto de las acciones presentes en las generaciones futuras y el medio ambiente. Su propuesta de una responsabilidad extendida requería una reflexión ética

que no solo abordara las consecuencias inmediatas, sino también los efectos a largo plazo. En su obra, Jonas formuló el *imperativo de responsabilidad*: «Actúa de tal manera que los efectos de tu acción sean compatibles con la permanencia de la vida humana genuina en la Tierra». Este imperativo, profundamente relevante en nuestra era tecnológica, enfatiza la obligación moral de salvaguardar la dignidad, autonomía e integridad tanto de la vida presente como de la futura.

Jonas no solo alteró el enfoque ético tradicional, sino que revitalizó el pensamiento anticipatorio, alineándolo con el concepto aristotélico de *causa final* como una guía práctica y ética para la acción. Esta perspectiva propone un retorno a los principios fundamentales de previsión y responsabilidad, que son cruciales para la preservación de la vida humana en un contexto en el que lo natural y lo artificial se entrelazan cada vez más.

La noción de *anticipación responsable* que se propone implica tanto un actuar de manera ética, considerando las consecuencias futuras (ética consecuencialista), como las obligaciones morales (ética deontológica) de nuestras acciones, además de adoptar una postura proactiva que promueva el cuidado responsable (ética de la virtud). Este enfoque es esencial en ámbitos como la atención médica, la educación y, por extensión, en la gobernanza de la IA, donde la vulnerabilidad del bien común debe ser protegida.

La *estructura moral* de los seres humanos proporciona una visión de nuestra capacidad única como agentes reflexivos con la responsabilidad de tomar decisiones que tienen implicaciones éticas. Como destaca Adela Cortina (2013), la existencia humana es inherentemente dramática debido a la necesidad constante de tomar decisiones y justificar nuestros actos. Nuestra libertad, aunque es un privilegio, también impone la carga de la justificación moral. Esta dinámi-

ca de libertad, decisión y responsabilidad constituye el eje ético de nuestra acción, y nos hace responsables no solo por las decisiones de alcance inmediato, sino por los futuros que contribuimos a crear.

La *anticipación responsable* responde a la pregunta crítica de los tomadores de decisiones: «¿Qué debería hacer ahora?». Esta interrogante, central en el enfoque de Robert Rosen (1985), nos lleva a considerar un enfoque paradigmáticamente diferente para la toma de decisiones. Su desarrollo conceptual sobre sistemas *anticipatorios* proporciona una nueva comprensión de cómo los sistemas biológicos y sociales pueden tomar decisiones informadas por el futuro, a diferencia de los sistemas reactivos, que simplemente responden a estímulos pasados.

A través de *sistemas anticipatorios* a nivel de los procesos cognitivos, se genera información y conocimiento que alimenta la toma de decisiones. La anticipación explícita (Poli, 2010) es la capacidad consciente de incorporar o generar información sobre un momento posterior para actuar en consecuencia. Los supuestos anticipatorios son las piezas operativas concretas de este proceso. Su consideración permite explorar la sinergia entre ética, intención y futuros potenciales, enriqueciendo tanto nuestra comprensión teórica como la aplicación ética práctica. Pueden actuar como un puente entre la teoría ética y la práctica, ofreciendo un enfoque matizado para la *anticipación responsable* que combina la reflexión ética con el pensamiento orientado al futuro.

El proceso de usar el futuro en la toma de decisiones

Aunque el futuro en sí mismo no existe, lo usamos cada vez que estamos anticipando. Este es el significado abreviado de *uso del futuro* y abarca los muchos propósitos y

formas de anticipación, entre ellas, la preparación, la planificación, así como la exploración y creación de alternativas en el presente (Miller, 2018, p. 10).

Un *uso del futuro* alfabetizado requiere una comprensión sobre cómo usamos aquello no existente (el futuro) para generar conocimiento y para informar la toma de decisiones. Aquí, comenzar por reconocer los sistemas y supuestos anticipatorios y cómo influyen en la manera que percibimos el presente, se vuelve un aspecto esencial. Al mismo tiempo, si reconocemos la naturaleza contingente del futuro —modelado por una multitud de posibles eventos y decisiones en creación permanente—, podemos entender mejor y relacionarnos con la incertidumbre y la complejidad.

Esto implica no solo imaginar una gama de futuros potenciales —como objetivos o escenarios adelante en el tiempo, es decir, futuros sustantivados— y evaluarlos en términos de deseabilidad y viabilidad, sino también el reconocimiento y selección de diferentes sistemas y supuestos anticipatorios (modelos subyacentes) que participan en esa percepción y la modelan, por lo que deben estar alineados con el propósito y con la naturaleza misma de los fenómenos y problemas.

En otras palabras, se trata de comprender y usar de manera diestra los sistemas y modelos que permiten incorporar lo no existente aún, a través de la reflexividad sobre los modos epistémicos en que esto se hace, ya que esto cambia lo que vemos en el presente, las oportunidades y desafíos que percibimos (que puede estar sesgado, incompleto o equivocado). Como resultado, la toma de decisiones se vuelve más informada, matizada y alineada con consideraciones de valor éticas, porque hay una mayor destreza para incorporar el futuro en el análisis.

Si bien esto puede resultar muy abstracto así planteado, lo es, porque nos estamos

refiriendo a procesos de alta cognición que hacen posible la anticipación misma (que es una acción con implicaciones prácticas). Además, la anticipación es esencialmente contrafactual, ya que incide antes y puede cambiar aquello a suceder, un aspecto más que fundamenta la necesidad de otras lógicas, metodologías y destrezas.

La *alfabetización en futuros* se convierte en una habilidad clave en este contexto y podemos equipararla a cualquier otro tipo de alfabetización (alfanumérica, computacional, emocional). Esta capacidad permite a individuos y organizaciones introducir el futuro más allá de la palabra *futuro* o de proyecciones o extrapolaciones del pasado (que es lo que suele hacerse), y en cambio, propiciar la reflexividad y creatividad necesarias para navegar la incertidumbre con agilidad y responsabilidad, asegurando que sus acciones presentes estén mejor informadas.

En la práctica, este enfoque transforma la forma en que se formulan las políticas, permitiendo que el uso del futuro se convierta en una herramienta poderosa para la gobernanza anticipatoria y la *anticipación responsable*.

Siguiendo el trabajo de Sripada (2016), podemos distinguir entre las fases de construcción y selección en los procesos de toma de decisiones. La fase de construcción implica la creación de opciones significativas basadas en la imaginación y la prospección de posibilidades futuras. Este proceso es fundamental para ampliar el conjunto de alternativas posibles, enriqueciendo así la toma de decisiones. La fase de selección, por su parte, implica evaluar y asignar valor a las opciones generadas, asegurando que las decisiones finales estén alineadas con principios éticos y futuros deseables.

De este modo, la *anticipación responsable* no se limita a la mera predicción o a la responsabilidad por tareas específicas, sino que

implica una actitud cuidadosa y reflexiva que atraviesa todo el proceso de toma de decisiones. Desde la formulación del problema y su replanteo, hasta una comprensión más profunda que se expresa en la capacidad de diversificar alternativas y elegir mejores opciones con capacidad de transformar antes, la *anticipación responsable* guía a los tomadores de decisiones a actuar con un sentido profundo de responsabilidad ética. Este enfoque garantiza que las decisiones presentes contribuyan a la creación de futuros deseables, mientras se minimizan los riesgos inherentes a la incertidumbre.

Consideraciones aplicadas para una gobernanza anticipatoria de la IA

Los Parlamentos son quizás las instituciones con mayor intensidad en el uso del futuro y, por lo tanto, responsabilidad con las sociedades y la humanidad en general, en términos de proporcionar garantías para una evolución segura de la IA: IA responsable, IA para el bien común.

A través de sus funciones principales, como la rendición de cuentas, supervisión, representación y legislación, tienen una influencia directa y concreta en las directrices para el desarrollo de la IA. Por este motivo, el Parlamento del Uruguay creó una Comisión Especial de Futuros como propuesta innovadora que le permita recorrer una curva de aprendizaje para la gobernanza anticipatoria. Como espacio plural de interrelación con su ciudadanía y con los demás ámbitos y niveles de gobierno, es una gran oportunidad para propulsar un ecosistema de gobernanza anticipatoria.

En este caso aplicado sobre gobernanza de la IA, el imperativo de una práctica de *anticipación responsable* es una condición *sine qua non*.

« Se trata de comprender y usar de manera diestra los sistemas y modelos que permiten incorporar lo no existente aún,»

Algunas recomendaciones para la gobernanza anticipatoria de la IA

Estas recomendaciones surgieron en el marco de la Segunda Cumbre Mundial de Comisiones de Futuros en Parlamentos, reunida en Montevideo, Uruguay, en 2023.

1. *Desarrollo de un marco de gobernanza anticipatorio de la IA.* Es crucial establecer un marco regulatorio global articulado con directivas a nivel internacional y regional. Este marco debe promover la cooperación internacional, asegurar el uso ético y responsable de la IA y regular su evolución para mitigar riesgos y maximizar beneficios.

2. *Promoción de la transparencia y explicabilidad de los algoritmos.* Implementar marcos que exijan la transparencia en el desarrollo de los algoritmos, para garantizar que las decisiones tomadas por los sistemas de IA sean comprensibles y auditables. La explicabilidad es crucial para evitar sesgos y fomentar la confianza pública. Se deben desarrollar sistemas de auditoría continua para monitorear el comportamiento de la IA. En la promoción de la transparencia y explicabilidad de los algoritmos, se puede destacar el caso de la Unión Europea, que ha establecido directrices claras para garantizar la responsabilidad y auditabilidad de los sistemas de IA, marcando un estándar global en el ámbito de la regulación tecnológica.

3. *Fomento de una gobernanza inclusiva y participativa.* Es esencial diseñar una gobernanza de IA que promueva la inclusión

de diversas partes interesadas (gobiernos, sector privado, sociedad civil, academia) y asegure que la tecnología beneficie a toda la sociedad. Se deben priorizar políticas que garanticen el acceso equitativo a las tecnologías emergentes, evitando que las brechas tecnológicas perpetúen desigualdades.

4. *Fortalecimiento de capacidades anticipatorias.* Los Parlamentos deben desarrollar capacidades anticipatorias para gestionar la evolución de la IA y prepararse para los cambios disruptivos. Esto incluye la creación de unidades especializadas en el *uso del futuro* dentro de los Parlamentos y el desarrollo de programas de capacitación para legisladores en temas de futuros, IA y complejidad. Esto es deseable que se extienda de manera articulada a otros espacios de gobierno y de la sociedad a modo de ecosistema de gobernanza anticipatoria.

5. *Uso de sandboxes regulatorios.* Implementar entornos de experimentación controlados (*sandboxes*) para probar y ajustar las regulaciones de la IA de manera iterativa. Esto permitirá a los Parlamentos adaptarse flexiblemente a los cambios tecnológicos, garantizando que las normativas evolucionen junto con la tecnología.

6. *Adopción de principios éticos fundamentales.* Incluir principios éticos en la gobernanza de la IA, tales como la transparencia, la equidad, la privacidad y la seguridad. Estos principios deben aplicarse a lo largo de todo el ciclo de vida de los sistemas de IA, desde su diseño hasta su implementación y uso.

7. *Promoción de la educación y la alfabetización en IA.* Desarrollar programas de educación y capacitación en IA para parlamentarios, ciudadanos y profesionales de distintos sectores. Esto fomentará una mayor comprensión de los riesgos y oportunidades de la IA, preparando a la socie-

« La alfabetización en futuros, la ética de la anticipación y el desarrollo de capacidades anticipatorias en los tomadores de decisión son esenciales para asegurar que la IA se desarrolle para el bien común. »

dad para participar de manera activa en los procesos de gobernanza.

8. *Cooperación internacional en gobernanza tecnológica.* Fomentar la cooperación internacional y el intercambio de mejores prácticas entre países para abordar los desafíos globales de la IA, promoviendo soluciones compartidas que puedan ser aplicadas a nivel local y global.

Estas recomendaciones buscan fortalecer la capacidad de los Parlamentos y otras instituciones de gobierno, de la academia, los desarrolladores y la sociedad civil en la gestión anticipatoria de la IA, y promover una gobernanza anticipatoria, participativa y basada en principios éticos que garantice el uso seguro y beneficioso de esta tecnología para el bien común, que incluye lo social y medioambiental.

Hacia una gobernanza anticipatoria responsable

La gobernanza anticipatoria de la IA es una necesidad fundamental en un contexto de cambios tecnológicos acelerados y profundos. El desarrollo de la AGI y la eventual aparición de la ASI plantean desafíos exis-

tenciales que no pueden ser abordados con enfoques tradicionales de gobernanza. La *anticipación responsable*, basada en la ética y la reflexividad, debe guiar el diseño de marcos regulatorios flexibles y colaborativos que permitan gestionar tanto los riesgos como las oportunidades que presenta la evolución de la IA.

El futuro de la IA no está escrito, y depende de las decisiones que tomemos hoy. La alfabetización en futuros, la ética de la anticipación y el desarrollo de capacidades anticipatorias en los tomadores de decisión son esenciales para asegurar que la IA se desarrolle para el bien común, garantizando que sus beneficios alcancen a toda la sociedad sin comprometer su seguridad o dignidad, asegurando que la IA evolucione de manera beneficiosa y equitativa para toda la humanidad. ♦

Referencias bibliográficas

- ARENDETT, H. (2008) [1958]. *La condición humana*. Barcelona: Paidós.
- CORTINA, A. (2013). *¿Para qué sirve realmente la ética?* Madrid: Paidós.
- GARRIDO, L. (2024). Responsible Anticipation. Futures literacy capacities to enhance ethical stance in anticipatory governance decision-making. Learnings and applications in Parliaments. En T. FULLER et al. (ed). *Towards Principles for Responsible Futures*. Lincoln University, Taylor and Francis (en imprenta).
- GLENN, J., y GARRIDO, L. (2023). *Parliaments and Artificial General Intelligence (AGI). An Anticipatory Governance Challenge*. idea Internacional.
- JONAS, H. (2014) [1984]. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago: University of Chicago Press.

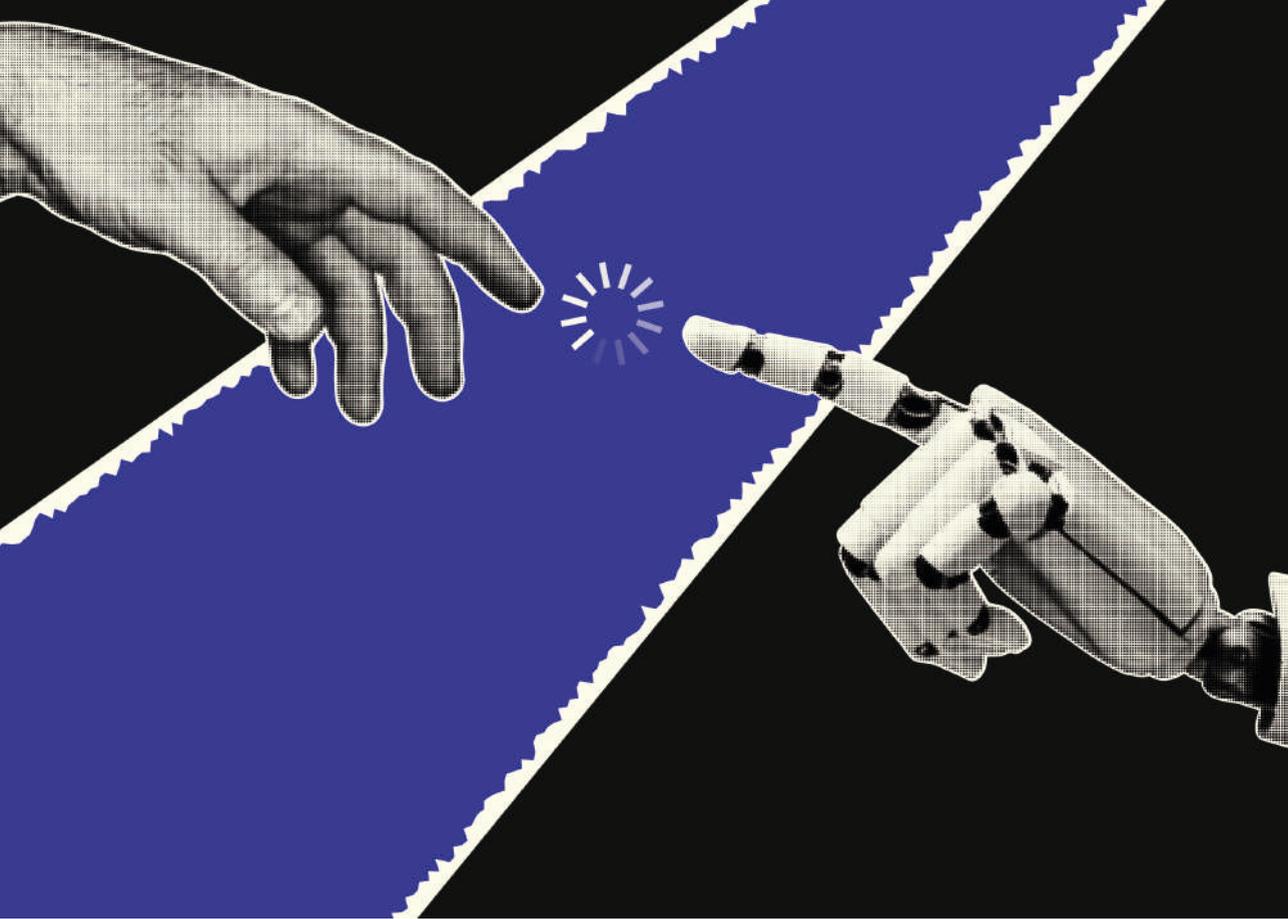
- MILLER, R. (2018). *Transforming the Future: Anticipation in the XXI Century*. París: Unesco. Nueva York: Routledge.
- MILLER, R., y POLI, R. (2010). Anticipatory Systems and the Philosophical Foundations of Futures Studies. *Foresight*, 12(3), 3-6.
- POLI, R. (2010). An Introduction to the Ontology of Anticipation. *Futures*, 42(7), 769-776.
- ROSEN, R. (1985). *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*. Oxford: Pergamon Press.
- RUSSELL, S., PERSET, K., y GROBELNIK, M. (2023). Updates to the OECD's definition of an AI system explained. *OECD.AI Policy Observatory*. <https://oecd.ai/en/wonk/ai-system-definition-update>
- SRIPADA, CH. (2016). Free Will and the Construction of Options. En M. SELIGMAN, P. RAILTON, R. BAUMEISTER y CH. SRIPADA (eds.), *Homo Prospectus*. Nueva York: Oxford University Press.



Lydia Garrido Luzardo

Antropóloga y futurista. Doctora en pensamiento complejo con maestría en investigación integrativa. Directora de la Cátedra

Unesco. en Anticipación Sociocultural y Resiliencia en el Instituto Sudamericano para Estudios sobre Resiliencia y Sostenibilidad del Uruguay. Asesora de la Comisión Especial de Futuros del Parlamento del Uruguay. LI: lydia-garrido-luzardo-aa486831



¿Es posible regular la IA? Experiencias a nivel global

XIMENA DOCARMO

La transición entre la inteligencia artificial capaz de interactuar en lenguaje natural y la llegada de ChatGPT 3.5 deja en evidencia las disparidades globales. Estas se expresan también en el acceso a las nuevas tecnologías. La necesidad de regular presenta desafíos que resulta indispensable encarar. Este artículo explica los modelos normativos posibles.

El dilema de control en la IA

Para el momento del lanzamiento de ChatGPT 3.5 a finales del 2022, en algunas regiones como la Unión Europea ya desde 2021 se habían iniciado debates públicos sobre promover una regulación ética de la IA, considerando los rápidos avances de esta tecnología. En contraste, a dos años del acceso público y masivo a la IA generativa, en diversas partes del mundo, el entendimiento y los esfuerzos para regular esta tecnología siguen siendo limitados.

Las razones de esta disparidad pueden ser múltiples. Para entender este fenómeno, resultan útiles las reflexiones de David Collingridge, quien en los años ochenta planteó el *dilema de control*. Este señala que «intentar controlar una tecnología es difícil, y no pocas veces imposible, porque durante sus primeras etapas, cuando se puede controlar, no se sabe lo suficiente sobre sus consecuencias sociales perjudiciales como para justificar el control de su desarrollo; pero para cuando estas consecuencias son evidentes, el control se ha vuelto costoso y lento».

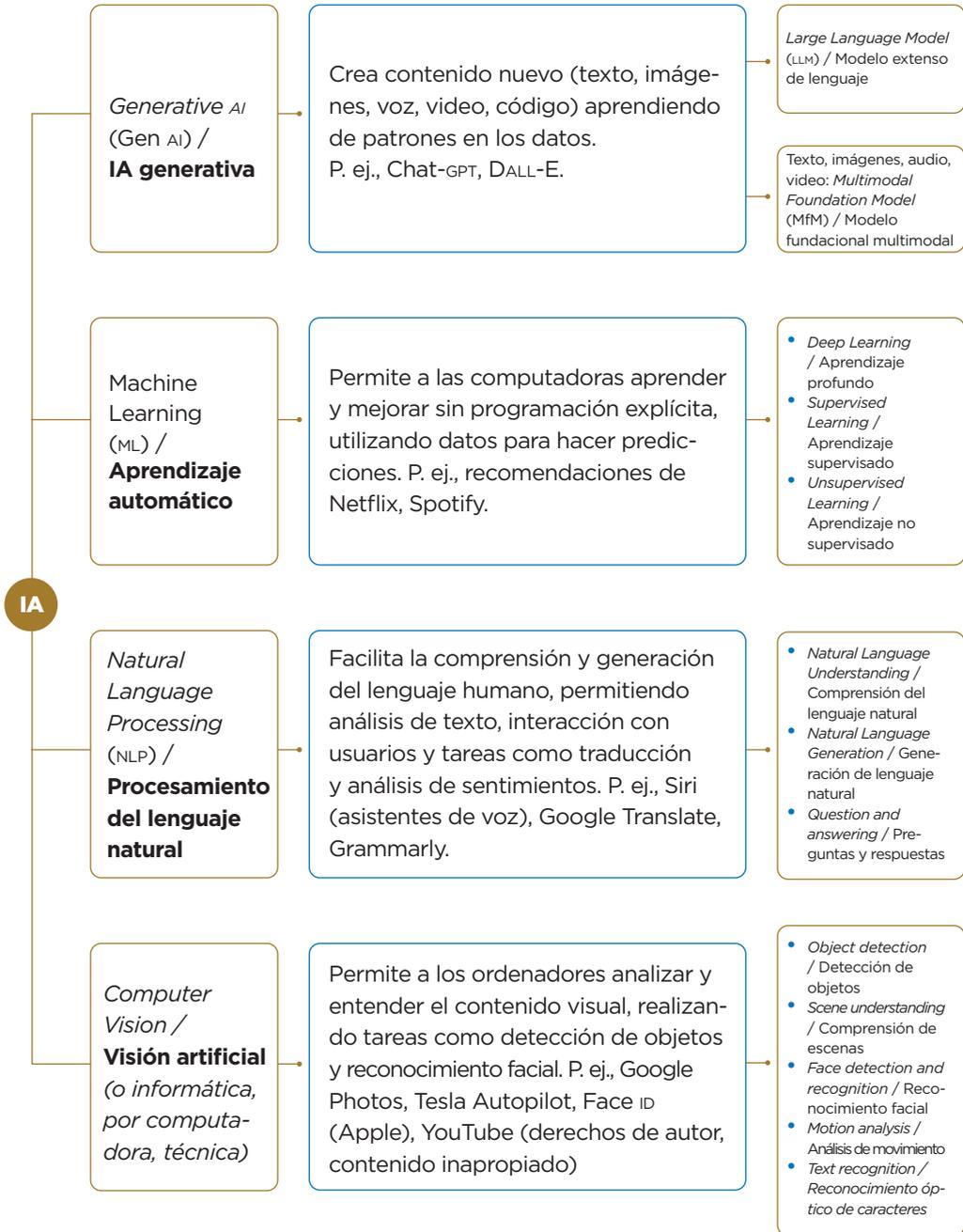
Al respecto, Collingridge destaca que, para prevenir efectos no deseados de una tecnología, deben cumplirse dos condiciones: primera, que se conozcan los efectos negativos que la tecnología podría causar, y

segunda, que sea posible modificarla para evitar dichos efectos. Dos condiciones que, en el contexto de la IA, parecen imposibles de prever en su magnitud real, lo que pondría en peligro nuestra capacidad para gestionar sus efectos de manera efectiva y oportuna.

¿Por qué regular la IA?

Antes de que ChatGPT 3.5 se convirtiera en un hito de los avances tecnológicos en 2022, la IA ya había comenzado a ser parte de nuestras vidas. La IA no es algo del futuro, es lo que despierta con nosotros cada mañana cuando miramos la pantalla de nuestro teléfono u otro aparato electrónico (ver cuadro 1). Desde el momento en que desbloqueamos nuestro *smartphone* recibimos sugerencias personalizadas sobre el clima, la música ideal para empezar el día, la ruta más rápida hacia el trabajo o incluso posibles respuestas para un mensaje de WhatsApp. Todo esto gracias a una rama de la IA llamada *machine learning* ‘aprendizaje automático’. El día sigue y al usar Face ID para desbloquear el teléfono o Google Lens para traducir una indicación en otro idioma o buscar información a partir de una imagen, estamos interactuando con otra rama de la IA: *computer vision* ‘visión artificial’, también

Cuadro 1. Ramas especializadas de la IA



Fuente: Elaboración y ejemplos propios. Adaptado de *A common understanding: simplified AI definitions from leading standards* (NSW Government, 2024).

llamada visión informática, visión por computadora o visión técnica). Esta interacción puede ser menos palpable; por ejemplo, al ver videos en YouTube, activamos otra pieza de esta tecnología para detectar contenido inapropiado o infracciones de derechos de autor en videos subidos a la plataforma.

La IA no solo está presente en nuestra vida personal, sino también en la profesional. Herramientas como ChatGPT, Gemini o Canva han transformado la manera en que trabajamos. Estas plataformas, basadas en IA generativa y procesamiento del lenguaje natural, nos permiten simplificar tareas complejas. Desde pedirle ayuda a Alexa o Siri, hasta escribir un correo electrónico en otro idioma con Google Translate o corregir textos con Grammarly, los programas potenciados por IA, aunque casi imperceptibles, son para muchas personas parte del día a día.

Pese a que la principal limitación de la IA radica en la necesidad de acceder a internet para su uso, el acceso es cada vez más masivo a nivel global. Según Naciones Unidas, al 2023 más del 65% de la población global estaba conectada a internet y más del 75% contaba con un teléfono móvil, una cifra que en diez años escalaría al 78%. Con tres de cuatro personas alrededor del mundo que antes de la llegada de ChatGPT 3.5 en 2022 comenzaban su día accediendo de una u otra forma a la IA, resulta lógico cuestionarse ¿qué ha cambiado? y ¿por qué se ha intensificado la discusión sobre la regulación de la IA?

El *dilema de control* explicado previamente puede orientar la respuesta a este tipo de preguntas. En las primeras etapas del desarrollo de la IA no se sabía lo suficiente sobre sus consecuencias. Hoy, los efectos de la IA, especialmente de la IA generativa, son cada vez más evidentes, lo que hace más necesaria su regulación, aunque eventualmente pueda llegar tardíamente y los costos asociados sean altos para la sociedad.

La IA representa un avance significativo en el desarrollo científico, con el potencial de cerrar brechas en sectores clave como la educación o la salud y de estimular la economía a través de la innovación. Pero también plantea serios desafíos para los derechos de las personas. Los modelos de IA, entrenados con información proveída por seres humanos, reflejan los defectos y sesgos de nuestra sociedad. La IA puede amplificar dichos sesgos reforzando la discriminación a determinados sectores de la población, facilitar el uso indebido de datos personales y la vulneración de la libertad de expresión, amplificar la desinformación, entre muchos otros efectos negativos. Por lo tanto, el equilibrio entre maximizar los beneficios de la IA y mitigar sus riesgos exige un enfoque ético en su regulación que pondere estos potenciales efectos.

Aunque las personas, de manera consentida o no, comparten diversos datos sobre sus preferencias, la omnipresencia de la IA, combinada con una limitada alfabetización tecnológica en IA, hace que los usuarios solo tengan un control parcial sobre el uso y la privacidad de sus datos. Los datos recopilados por empresas privadas, como la trazabilidad de las compras, búsquedas en línea o interacciones en redes sociales, así como aquellos obtenidos por empleadores o incluso gobiernos, pueden ser utilizados para diferentes fines. En contextos autoritarios, esta capacidad de vigilancia y control puede tener implicaciones preocupantes, exacerbando riesgos para los derechos individuales y las libertades fundamentales.

» La omnipresencia de la IA [...] hace que los usuarios solo tengan un control parcial sobre el uso y la privacidad de sus datos. «

Ante el inminente uso de la información de las personas, es crucial que los gobiernos aborden cuestiones clave sobre la privacidad, la seguridad y el uso ético tanto de los datos introducidos como de los resultados procesados por la IA. Sin marcos regulatorios sólidos, el riesgo de un uso irresponsable de la IA aumenta considerablemente. Dicha regulación debería incluir mecanismos que aseguren un uso responsable y sostenible de la tecnología, así como fomentar la alfabetización en IA para mitigar los riesgos.

IA en la UE (2021-2030)

Con el objetivo de garantizar que los sistemas de IA usados en la Unión Europea sean «seguros, transparentes, trazables, no discriminatorios y respetuosos con el medio ambiente», con 523 votos a favor en 705 escaños, y tras tres años de debates, en marzo de 2024 el Parlamento Europeo adoptó formalmente la primera ley que regula la IA: la Ley de IA de la Unión Europea (UE). La legislación, originalmente presentada por la

Cuadro 2. Categorías de tecnologías de IA



Fuente: Elaboración propia. Adaptado de European Parliament (2023).

Comisión Europea en abril de 2021, proponía establecer el primer marco regulatorio para la IA en esta región.

Actualmente no existe una definición unánime global para la IA, por lo cual cada país o región que la regula va a ponderar diversos elementos o categorías en el proceso de reglamentar su uso. En el caso de la UE (cuadro 2), la ley de IA se ha enfocado en la IA de propósito general (GPAI, por sus siglas en inglés); los también llamados modelos fundacionales tienen capacidades generativas y están diseñados para realizar una amplia gama de tareas inteligentes. Como parte de la inteligencia artificial general (*artificial general intelligence*, AGI) la existencia de la GPAI es posible gracias a los modelos extensos de lenguaje (LLM) y su capacidad generativa. En contraste, la inteligencia artificial estrecha (*artificial narrow intelligence*, ANI) únicamente puede realizar tareas específicas y predefinidas. Estos tres elementos (GPAI, AGI y ANI) resultan cruciales para comprender la clasificación de riesgo de la IA para la UE y su respectivo cronograma de implementación.

La ley clasifica los sistemas de IA en cuatro categorías: riesgo inaceptable, alto, limitado y mínimo, considerando acciones concretas y dando flexibilidad para adoptar modificaciones si determinados usos evolucionan generando un mayor riesgo:

» No existe una definición unánime global para la IA, por lo cual cada país o región que la regula va a ponderar diversos elementos o categorías en el proceso de reglamentar su uso. «

- *Riesgo inaceptable*. Están completamente prohibidos los sistemas destinados a manipular el comportamiento humano o la clasificación social, incluyendo aquellos que emplean técnicas manipulativas, explotan vulnerabilidades de personas en situaciones desfavorecidas, y los que implementan clasificación social.
- *Riesgo alto*. Están sujetos a estrictas normas de seguridad, transparencia y supervisión humana los sistemas que tienen un impacto significativo en la seguridad o los derechos fundamentales, como los utilizados en la educación, empleo, infraestructuras críticas y la administración de justicia.
- *Riesgo limitado*. Tienen obligaciones menos estrictas, como asegurar que los usuarios sepan que están interactuando con una IA, sistemas como los chatbots o *deepfakes*.
- *Riesgo mínimo*. Están exentos de regulaciones más estrictas los videojuegos habilitados por IA o los filtros de spam. Esto podría cambiar con el avance de la IA generativa.

La ley *restringe* el uso de reconocimiento facial en tiempo real en espacios públicos, salvo excepciones como la búsqueda de personas desaparecidas o la prevención de amenazas terroristas.

Esta norma afecta a 27 Estados miembros. Dada la complejidad del sistema de la Unión Europea, la Ley de IA publicada en julio de 2024 y con entrada en vigor en agosto del mismo año tiene previsto un largo proceso de implementación con un cronograma escalonado para la aplicación de las diferentes obligaciones (cuadro 3). Mientras que todos los Estados miembros están llamados a informar hasta noviembre de 2024 sobre las autoridades responsables de la implementación de la normativa, las

Cuadro 3. Calendario de aplicación de la Ley de IA de la UE

12.7.2024	Se publica la Ley de AI de la UE en el Diario Oficial de la UE el 12 de julio de 2024.
1.8.2024	La Ley de AI de la UE entra en vigor el 1 de agosto de 2024 (art. 113).
2.2.2025	Las normas sobre objeto, ámbito de aplicación, definiciones, alfabetización en materia de IA y prohibiciones entran en vigor el 2 de febrero de 2025 (art. 113.a).
2.8.2025	Las normas sobre notificaciones, modelos GPAI, determinadas cuestiones de aplicación y sanciones entran en vigor el 2 de agosto de 2025 (art. 113.b).
2.8.2026	La Ley de AI de la UE se aplica, el periodo de gracia general para los sistemas de AI de alto riesgo finaliza y la mayor parte de las disposiciones operativas entran en vigor el 2 de agosto de 2026 (arts. 111.2 y 113).
2.8.2027	Normas sobre IA de alto riesgo en virtud del apartado 1 del artículo 6 entran en vigor el 2 de agosto de 2027 (art. 113.c).
2.8.2030	El período de gracia para sistemas de IA de alto riesgo destinados a ser utilizados por autoridades públicas finaliza el 2 de agosto de 2030 (art. 111.2).

Fuente: Extracto del EU AI Act enforcement timeline (2024) y White & Case (2024), versión extendida en inglés.

prohibiciones a ciertos sistemas de IA de riesgo inaceptable entrarán en vigor seis meses después de la promulgación de la ley, a partir de febrero de 2025. Asimismo, las regulaciones para los modelos GPAI se implementarán en 12 meses, y los sistemas de alto riesgo comenzarán a ser regulados en 24 y 36 meses. En este periodo, progresivamente entrarán en vigor otras normas claves sobre gobernanza, confidencialidad y sanciones.

A nivel temporal, la Ley de IA de la UE sigue un enfoque gradual que comienza con la alfabetización y prohibiciones, seguido de la introducción de normas sobre notificaciones, modelos de IA y sanciones. A medida que avanza, finaliza el periodo de gracia para los sistemas de IA de alto riesgo, se aplican disposiciones operativas y se implementan nor-

mas específicas para la IA de alto riesgo. Para garantizar la implementación de la norma, se ha previsto la creación de la Oficina de IA de la UE dentro de la Comisión Europea.

Modelos de regulación en desarrollo

En contraste con el enfoque adoptado por la Unión Europea, otras regiones del mundo con un alto grado de desarrollo en IA priorizan distintos componentes en sus modelos regulatorios y de adopción tecnológica. En el libro *Digital Empires: The Global Battle to Regulate Technology*, Anu Bradford reflexiona sobre los contrastes entre el modelo europeo centrado en establecer estándares de regulación a nivel global, el modelo *estadounidense* que fomenta al



sector privado, y el *modelo chino* impulsado por el uso de los recursos estatales.

Así explica que en el *modelo estadounidense*, centrado en el mercado, el papel gubernamental es limitado y permite que las grandes empresas tecnológicas lideren la gobernanza. Fomenta un entorno propicio a través de incentivos para la innovación y exporta su influencia a través de los servicios y tecnologías, consolidando su poder privado en la economía global.

En cuanto a la regulación de la IA, Estados Unidos carece de una legislación federal integral, su enfoque cuenta con leyes y directrices de alcance limitado. Entre las normas claves destaca la Ley de Iniciativa Nacional de IA (*National AI Initiative Act*) de 2020, centrada en el fomento de la investigación y desarrollo en este campo, junto con la creación de la Oficina de la Iniciativa Nacional de Inteligencia Artificial (*National Artificial Intelligence Initiative Office*), encargada de implementar la estrategia nacional. En octubre de 2023, la Casa Blanca emitió la Orden Ejecutiva sobre el Desarrollo y la Utilización Seguros y Fiables de la IA, que

establece directrices para el desarrollo de estándares federales incluyendo elementos de transparencia en los resultados de las pruebas de seguridad. En el último año, diversos estados han liderado iniciativas sobre la regulación de los sistemas de alto riesgo, la discriminación algorítmica o la toma de decisiones automatizadas. La tendencia general es hacia una mayor regulación sectorial y en los estados; no obstante, se prevé la continuación de discusiones públicas tanto sobre la implementación de una regulación cohesiva de la IA como la creación de una autoridad federal.

Mientras tanto, el *modelo chino*, impulsado por el Estado, busca establecer al país como una superpotencia tecnológica mediante la utilización de recursos estatales. Este modelo se manifiesta en la vigilancia, censura y propaganda, y acciones dirigidas a preservar el control político. A su vez, China exporta su poder de infraestructura, desarrollando redes 5G, centros de datos y ciudades inteligentes.

A nivel regulatorio, en 2023 fue publicada la primera regulación administrativa

específica sobre IA generativa, denominada *Medidas provisionales para la gestión de los servicios de IA generativa*. En esta no se categorizan los riesgos, pero ciertos servicios, como los que cuentan con «atributos de opinión pública o capacidad de movilización social» están sujetos a un escrutinio más riguroso, incluyendo evaluaciones de seguridad, y a requisitos de aplicación general como la moderación y el etiquetado de contenidos. Entre las normas de etiqueta se destacan requisitos como la defensa de los valores socialistas o no generar contenidos prohibidos que inciten a la subversión del poder estatal o el derrocamiento del sistema socialista, que pongan en peligro la seguridad y los intereses nacionales. La responsabilidad de la regulación de la IA generativa recae principalmente en la *administración del ciberespacio de China*.

A esta discusión se suman otros gigantes en la escena global. Según el Foro Económico Mundial (WEF), las principales cinco economías más grandes del mundo han avanzado significativamente en el desarrollo de ecosistemas de IA. A esta lista, además de Estados Unidos, China y Alemania (miembro de la UE), se incorporan Japón e India. Aunque ambos carecen de leyes específicas sobre IA, han adoptado enfoques distintos para abordar su regulación.

Japón a nivel global ha sido actor clave del lanzamiento del *Código de conducta para organizaciones que desarrollan sistemas avanzados de IA*, en el contexto del G7 en 2024, un instrumento que compila 11 recomendaciones, con un enfoque basado en riesgo. A nivel interno, el país sigue una estrategia de *soft law* (derecho blando), promoviendo la gobernanza de la IA a través de directrices que buscan minimizar riesgos mientras priorizan impulsar la innovación. En 2024 también fue publicada la *AI Guidelines for Business Version 1.0*, una directriz

❖ **La experiencia internacional en la regulación de la IA ofrece valiosas lecciones para América Latina, tanto en términos de promover la innovación como de proteger a las personas de los riesgos asociados a esta tecnología.** ❖

no vinculante que busca promover esfuerzos voluntarios siguiendo un enfoque basado en riesgos. Sin embargo, un reciente proyecto de Ley de IA podría redireccionar la estrategia actual a un enfoque de *hard law* (ley dura), incluyendo la supervisión a desarrolladores y la imposición de multas y sanciones en caso de incumplimiento.

Por su parte, India ha establecido marcos sectoriales, como en finanzas y salud, y su enfoque está guiado por la *Estrategia nacional de IA*, de 2018 y los *Principios operacionales para una IA responsable*, de 2021, que priorizan la capacitación e incentivos al diseño ético de IA. Aunque las regulaciones son bastante limitadas, se espera que la futura Ley de India Digital delimite y regule sistemas de IA de alto riesgo.

América Latina: ¿controles posibles?

La experiencia internacional en la regulación de la IA ofrece valiosas lecciones para América Latina, tanto en términos de promover la innovación como de proteger a las personas de los riesgos asociados a esta tecnología. Sin embargo, pensar en una re-

gulación unificada para la región puede no ser realista ni efectivo, dado que cada país avanza a su propio ritmo en la creación de ecosistemas para el desarrollo de IA.

Recomendaciones recientes, como la resolución publicada en 2024 por Naciones Unidas denominada *Aprovechar las oportunidades de sistemas seguros y fiables de IA para el desarrollo sostenible* y la *Recomendación sobre la IA* de la OCDE adoptada originalmente en 2019 y actualizada en 2021, subrayan la necesidad de una gobernanza clara, inversión en infraestructura tecnológica y educación en habilidades digitales. Pero, para los países de América Latina, estos desafíos se abordarán desde una perspectiva más flexible, considerando las distintas realidades y capacidades de cada nación. Chile, Brasil y Uruguay lideran en investigación y desarrollo de IA, según el índice latinoamericano de IA (ILIA) 2024, pero los avances no son homogéneos en toda la región.

En un contexto tan complejo y que avanza rápidamente, América Latina debe equilibrar la promoción de la IA con la protección de derechos fundamentales, aprovechando esta tecnología para un desarrollo inclusivo y sostenible. La clave será diseñar marcos regulatorios que permitan una implementación responsable, respetando la diversidad y el ritmo de evolución de cada país, sin comprometer los derechos de las personas ni, en última instancia, la democracia. ♦

Referencias

- COLLINGRIDGE, D. (1980). *The Social Control of Technology*. Nueva York: St. Martin's Press.
- EU ARTIFICIAL INTELLIGENCE ACT. (2024). <https://artificialintelligenceact.eu/>
- European Parliament. (2023). *General -purpose artificial intelligence*. <https://>

[www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA\(2023\)745708_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA(2023)745708_EN.pdf)

ILIA, ÍNDICE LATINOAMERICANO DE INTELIGENCIA ARTIFICIAL. (2024). <https://indicelatam.cl/>

Naciones Unidas, Asuntos económicos. (2023). Más del 75% de la población mundial tiene un teléfono celular y más del 65% usa el internet. <https://news.un.org/es/story/2023/12/1526712>

NETFLIX. (2024). *Y ahora qué. El futuro según Bill Gates*. <https://www.netflix.com/us-es/title/81609333>

Parlamento Europeo, Ley de IA de la UE: primera normativa sobre inteligencia artificial (2024) <https://www.europarl.europa.eu/topics/es/article/20230601STO93804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial>

WHITE & CASE. (2024). The global dash to regulate AI. <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker#articles>

WORLD ECONOMIC FORUM. (2024, 2 de junio). Así es como los capitalistas de riesgo invierten en la IA en cinco países. <https://es.weforum.org/agenda/2024/06/como-esta-invirtiend-el-capital-riesgo-en-ia-en-las-cinco-principales-economias-mundiales-y-configurand-el-ecosistema-de-la-ia/>



Ximena Docarmo

Fundadora de Innovalab, entrenadora política y máster en políticas públicas por la Hertie School of Governance, Berlín.

X: @xidocarmo

IG: @xidocarmo

Fb: xidocarmo

LI: xidocarmo



De las redes masivas al voto personalizado

JESÚS DELGADO VALERY

El uso de inteligencia artificial transforma las campañas electorales con herramientas como el *microtargeting*. Las campañas pueden personalizar su comunicación política e influir en las decisiones de los votantes. Esto ofrece nuevas oportunidades y plantea desafíos éticos relacionados con la desinformación y el uso de datos personales, aspectos que impactan en la calidad de la democracia.

Durante los últimos años, el uso de las inteligencias artificiales se ha vuelto cada vez más común, especialmente en el campo de la generación de imágenes. La dinámica de los partidos políticos y los procesos electorales han sido notablemente afectados por este fenómeno. La masificación de las redes sociales fue sucedida por la irrupción de estas nuevas tecnologías, que juegan un papel cada vez más sensible en la difusión de propaganda política. Esto ha planteado un sinnúmero de debates, a medida que se desarrollan nuevas y más sofisticadas modalidades, en cuanto a planteos éticos y técnicos.

En principio, la generación de contenido por medio de la IA permite la creación de imágenes basadas en la solicitud del usuario. Esto da pie a que una campaña pueda construir un imaginario que represente un futuro eventual (como una forma de mostrar el resultado positivo de la aplicación de determinadas políticas o el negativo de otras). Sin embargo, esto ya desde el vamos trae aparejado el debate sobre la creación artificial de imágenes que no se condicen con la realidad, lo que puede entrar en la categoría de *fakenews* o narrativas falsas o sesgadas. El mismo resultado se produce con las inteligencias artificiales que permi-

ten emular la voz de una persona y que promueven la difusión de audios falsificados.

En los últimos años, muchas campañas han hecho uso de IA. Lo vemos en la superficie, para generar videos que muestran un escenario alternativo muy positivo en caso de ganar las elecciones (o, por el contrario, un escenario devastador en caso de una derrota), como hicieron las campañas de Sergio Massa y de Javier Milei en las elecciones argentinas de 2023, o las imágenes generadas por inteligencia artificial de Donald Trump siendo detenido después de su condena judicial (representando escenarios de arresto cada vez más inverosímiles, con el objetivo de generar viralización y memes).

Las redes sociales han equilibrado el acceso al electorado, en particular, a sectores que pudieran verse marginados de los principales medios de comunicación y que así pueden acceder a una vía de difusión de fácil manejo y adquisición barata. Sin embargo, su rápido alcance también permite la masificación de *fakenews* con mucha mayor facilidad y vuelve además más compleja su refutación (*factchecking*).

Aunque su aparición es reciente, ya existen muchos trabajos sobre el impacto de la inteligencia artificial en los procesos electorales, sobre todo en la generación de

contenido para campañas. En este trabajo intentaremos dar luz sobre el impacto de la inteligencia artificial en el diseño y armado de redes, así como en la organización del apoyo social, un tema menos estudiado.

IA en las campañas electorales

Se conoce como inteligencia artificial

[...] una disciplina perteneciente a las ciencias de la computación, que plantea modelos computacionales de aprendizaje basado en redes neuronales biológicas humanas. En este sentido, se han planteado diversos modelos de IA, que gracias a los avances en la tecnología computacional han permitido desarrollos de sistemas inteligentes que facilitan procesar mayor cantidad de datos en un tiempo menor, agilizando la toma de decisiones (Márquez Díaz, 2020).

La noción de inteligencia artificial se remonta a la segunda mitad del siglo xx. La famosa *prueba de Turing* planteaba el desafío de que personas comunes no pudieran diferenciar si estaban interactuando con una persona o con un *chatbot* en una conversación escrita (Turing, 1950). No obstante, el entusiasmo mundial por las IA no se alcanzó sino hasta hace pocos años. A partir de 2023 comenzaron a generarse imágenes con inteligencia artificial que podrían emular fotografías humanas.

En el plano de la política, su impacto ha sido notorio. Las campañas electorales se caracterizan por la necesidad de persuadir a los electores para movilizarlos a apoyar a determinada opción y, en los últimos años, también para desmovilizarlos o rechazar una opción, en caso de que atraerlos a la propia sea imposible. Si bien no es lo mismo

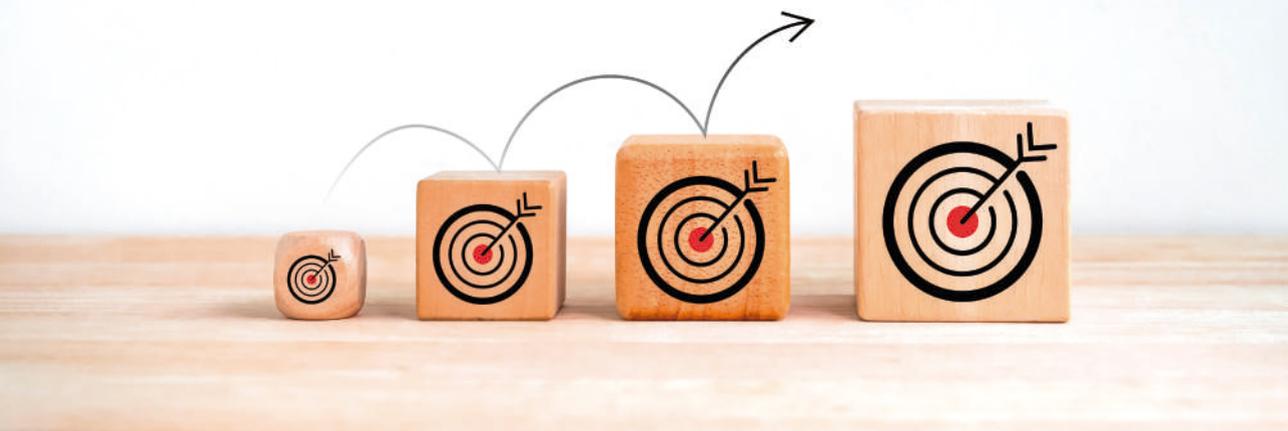
hacer campaña hoy que hace cincuenta o cien años, el objetivo sigue siendo el mismo.

Las campañas electorales del siglo xx se caracterizaban por ser masivas y abordar temas generales. Aunque hay distintas teorías sobre el comportamiento del voto, hay consenso en la literatura sobre la importancia que tenía la clase social en las decisiones de los electores, cuando este era un clivaje determinante.

Las campañas se hacían en gran medida a través de mítines populares de los principales líderes, y los cuadros de los partidos hacían su trabajo en las comunidades. Los canales de difusión masivos eran la radio, la televisión y la prensa escrita; y era una «característica común a todos ellos que son indiscriminados. Dicho con otras palabras, que no están adaptados para el perfil concreto, sino que buscaban convencer al conjunto de la sociedad» (Cebrián Beltrán, 2024).

En las campañas de las elecciones modernas o tradicionales la comunicación se daba en un contexto controlado, existían limitaciones de formato que impedían o reducían la agresividad y la desinformación; y en el electorado prevalecía un espíritu moderado que desalentaba y hasta castigaba el extremismo. Por otro lado, los medios de comunicación eran reducidos y su impacto y alcance dependían de su prestigio, de manera que eran muy cuidadosos en contrastar y comprobar la información que difundían (Rubio Núñez et al., 2024).

Sin embargo, la crisis de los grandes partidos del siglo xx en Europa y América Latina tuvo su correlato en la transformación del electorado, que pasó de ser una masa más o menos uniforme, a convertirse en numerosos subgrupos articulados alrededor de ideales mucho más específicos (ecologistas, defensores de la participación política de las mujeres, de las minorías, partidos regionalistas, separatistas-independentistas, etc.).



« El electorado pasó de ser una masa más o menos uniforme a convertirse en numerosos subgrupos articulados alrededor de ideales mucho más específicos. »

Esta nueva realidad obligó a las agrupaciones políticas a hacer esfuerzos para entender mejor a su electorado y usar las nuevas tecnologías, tanto para comprenderlo como para no dejar de representarlo.

La llegada de las redes sociales, por supuesto, ha resultado un aliciente revolucionario, quizás de la misma talla o mayor que la televisión. Aunque ninguno de los métodos usados anteriormente dejó de emplearse, las redes han implicado una masificación de la información y permitido llegar a casi todos los sectores de una sociedad en plazos mucho más cortos. Esto ha resultado un ba-

che para los sectores mayoritarios o incluso para regímenes autoritarios y su pretensión de mantener un control de la información.

Con las tecnologías actuales, incluso en un contexto autoritario, un candidato opositor puede diseñar una campaña coherente y competitiva que le permita el acceso a un amplio número de recursos (*spots*, *jingles*, entrevistas y difusión de propaganda, entre otros), algo imposible si se piensa en los contextos autoritarios del siglo xx.

Esto ocurre en las democracias. Hemos sido testigos de nuevas formas de hacer campaña que desestiman en gran medida la presencialidad y los actos masivos. Por ejemplo, en las elecciones presidenciales de Chile de 2021 se presentó como candidato Franco Parisi (Partido de la Gente), aunque su lugar de residencia era Estados Unidos. En medio de la pandemia, situación que promovió distintas herramientas de comunicación a distancia, Parisi hizo su campaña y obtuvo 900.000 votos, el equivalente al 12,8% de los sufragios (Serval, 2021).

Microtargeting: conocer al elector

Con estas nuevas tecnologías se puede entonces hacer algo que las anteriores no permitían o limitaban: crear mensajes políticos personalizados dirigidos a segmentos muy específicos de votantes, lo que se conoce como *microtargeting*. Esta estrategia se basa en la segmentación de la audiencia a la que se dirigirán los mensajes, mediante información demográfica, comportamental y de preferencias para diseñar mensajes que resuenen con las características y necesidades de grupos reducidos de personas, también conocidos como *microsegmentos* o *clústeres*. La IA brinda mecanismos sofisticados y novedosos para esta labor.

« Esta estrategia [el *microtargeting*] se basa en la segmentación de la audiencia a la que se dirigirán los mensajes, mediante información demográfica, comportamental y de preferencias para diseñar mensajes. »

El flujo es el siguiente: las personas dan sus datos personales al usar aplicaciones, visitar páginas webs, comprar por internet, etc. La suma de toda esta información, su análisis y tratamiento para correlacionar datos e identificar patrones y tendencias se conoce como *big data* o *macrodatos*. Esta información es la base con la que trabajan los algoritmos, que no son más que «una

secuencia de comandos para que una computadora convierta una entrada en una salida. Por ejemplo, una lista de personas según su edad. El ordenador toma las edades de las personas de la lista (entrada) y produce la nueva clasificación de la lista (salida)» (FRA, 2018).

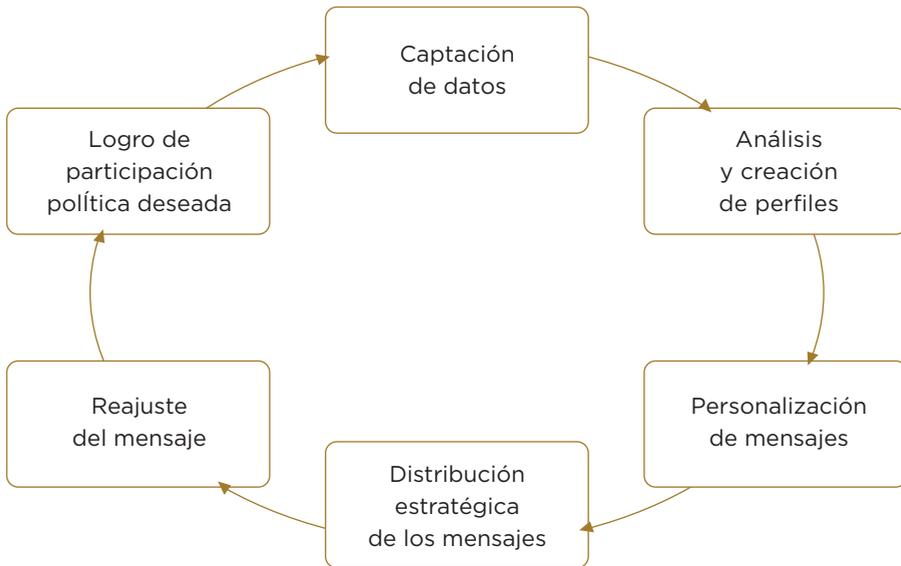
En materia electoral, lo que nos atañe en este artículo, el algoritmo aplicado a los macrodatos tendrá como objeto perfilar (*profiling*) a los electores, es decir, la información será analizada con el objetivo de comprender su postura política y predecir e influir sobre su comportamiento electoral. Este proceso termina con la conformación de los microsegmentos o clústeres, es decir, agrupaciones de individuos con perfiles similares.

Una vez caracterizados los distintos grupos, se diseñarán mensajes específicos para cada uno. El contenido de estos mensajes dependerá de los objetivos que se tengan. Por ejemplo, si se identifica un grupo de potenciales electores de un partido que están desmotivados o desmovilizados, se pueden diseñar mensajes que busquen involucrarlos en el proceso electoral, sensibilizarlos con la situación política y convencerlos de que su voto es necesario.

Por el contrario, si se identifica un segmento que tiende a votar por el adversario, se pueden diseñar mensajes específicos para desmovilizarlos, por ejemplo, con casos de corrupción de ese partido o declaraciones de sus líderes que vayan en contra de las convicciones de ese segmento.

Muchos de los mensajes diseñados no recurren a la capacidad analítica de los electores, sino a sus emociones. En este sentido, es cada vez más común encontrar temas polémicos (migración, aborto, religión) en la discusión política para exacerbar las emociones de los electores y simplificar el debate político. Esta tendencia ha

Ciclo de la inteligencia artificial en la participación política



Fuente: Cebrián Beltrán (2024)

promovido la difusión de desinformación como una herramienta para consolidar al electorado y el uso de la IA generativa para crear y manipular contenido audiovisual para fomentar la indignación.

Esta tecnología es tan vertiginosa que permite analizar su impacto en tiempo real e introducir cambios para hacer correcciones. Mientras que en las campañas modernas se debía acudir a los estudios de opinión tradicionales, como las encuestas o los grupos focales, para conocer el impacto de un mensaje, una propuesta o un eslogan, con el *microtargeting* y las redes sociales en cuestión de minutos se pueden hacer análisis preliminares sobre la efectividad de las estrategias usadas en la campaña.

Esta nueva estrategia permite que las fuerzas políticas y los candidatos accedan a las preocupaciones de determinados segmentos que antes podían pasar inadvertidas, a pesar de su elevado alcance electoral,

y puede aumentar la eficiencia de las campañas electorales por su capacidad de hacer análisis en tiempo real.

Sin embargo, el *microtargeting* también se utiliza para presentar a los electores información engañosa o sesgada, difundir *fakenews*, desinformación, para la generación de burbujas que limiten el debate, minando la confianza en las instituciones y las bases de la democracia.

Impacto en el resultado electoral

Aún no es fácil medir el impacto de las distintas herramientas provistas por la IA en los resultados electorales. Establecer una metodología rigurosa para tal fin resulta, al menos por el momento, imposible. Sin embargo, sí hay mecanismos para analizar el entorno comunicacional y registrar las discusiones que tiene una sociedad específica.

En ese sentido, es posible medir el éxito de las herramientas impulsadas por la IA en la instalación de temas en la agenda pública, así como en las posturas que toman los ciudadanos al respecto.

Puede asegurarse también que ha habido ciertos temas o ideas clave en algunas elecciones que han impactado de manera determinante en los resultados. Y que el uso de tecnologías como el *microtargeting* ha maximizado la influencia de estos temas en el electorado.

A continuación, se mencionan algunos casos emblemáticos en los que la tecnología, y en especial el uso de *bigdata* para diseñar campañas específicas, ha tenido impacto en el resultado electoral.

2008: el caso Obama

Mucho antes de que se hablara de IA hubo un cambio de paradigma que marcó un antes y un después en las campañas electorales: el *microtargeting*. En las elecciones de los Estados Unidos de 2008, el equipo del entonces candidato presidencial Barack Obama logró hacer un perfil de cada votante del país que se enfocaba en dos puntos: si iba a votar, y si lo haría por Obama. A partir de esta información se diseñaron estrategias para influir en sus decisiones y el candidato finalmente terminó llegando a la Casa Blanca. Fue una de las primeras experiencias masivas de perfilamiento de electores, aunque no con la sofisticación que conocemos actualmente.

2016: Cambridge Analytica

La elección de Donald Trump quedó marcada por el escándalo de Cambridge Analytica y Facebook, que se convertiría

en un caso de estudio sobre el uso de datos personales para diseñar mensajes políticos ultrasegmentados.

Aleksandr Kogan, profesor de la Universidad de Cambridge, diseñó un test de personalidad para Facebook en 2013 con el que obtuvo los datos de 50 millones de personas. El test fue efectivamente completado por 265.000 usuarios de la red social, que para hacerlo debían autorizar que se accediera a la información de su red de amigos, sin el permiso de estos.

Con los datos de estas personas, Cambridge Analytica creó perfiles psicológicos y diseñó mensajes específicos para influir en sus preferencias políticas, incluso difundiendo *fakenews* (BBC Mundo, 2018).

2023 en Argentina: imágenes y videos



En las elecciones presidenciales de Argentina de 2023, los equipos de los candidatos al balotaje, Sergio Massa y Javier Milei, recurrieron intensamente a la inteligencia

« La irrupción del *bigdata* y la IA representa un reto no solo para los partidos y el electorado, sino para las instituciones, la integridad del espacio comunicacional y finalmente el sistema democrático. »

artificial generativa para crear imágenes y videos de autopromoción y ataques al oponente. El equipo del candidato oficialista empleó la IA para crear afiches y videos que lo mostraban como un líder fuerte y carismático, inspirados en estilos soviéticos y la cultura pop. Por su parte, Milei respondió con imágenes que hacían alusión a un león *liberando a la Argentina* y a Massa como un líder comunista. El uso en la campaña de Massa de una IA generadora de videos para retratar un eventual futuro distópico en caso de un triunfo de Javier Milei fue objeto de controversia (Nicas y Cholakian Herrera, 2023).

2024 en India: *chatbots*

En India, en el contexto de la campaña para las elecciones generales de 2024 se desató una polémica sobre los *deepfakes* en redes sociales cuando un usuario le preguntó a la herramienta de IA de Google, Gemini, sobre el supuesto carácter fascista del primer ministro Narendra Modi y el Partido Popular Indio (BJP), y recibió como respuesta que el gobierno de Modi estaba «acusado de implementar políticas que algunos expertos han caracterizado como fascistas» (Dillon,

2024). El ministro de Estado indio para la Electrónica y la TI, Rajeev Chandrasekhar, criticó la respuesta, señalando que violaba las leyes del país. Este evento subraya la creciente preocupación en India sobre la desinformación y el uso de IA en el contexto electoral. Google reaccionó rápidamente afirmando que estaba «trabajando» para «mejorar la fiabilidad» de la herramienta (Mukherjee, 2024).

Ejemplos en Dinamarca y Reino Unido

El Synthetic Party es un partido político en Dinamarca dirigido por una IA llamada Leader Lars, un *chatbot* accesible a través de Discord. Su objetivo es atraer a los ciudadanos que no suelen votar y colocar la tecnología en el centro del debate político, promoviendo la convivencia entre IA y personas, así como la regulación de la responsabilidad de la IA. El partido, que se define como sintético, desarrolla su programa basado en las propuestas de partidos minoritarios daneses desde 1970. Aunque liderado por una IA, el proyecto es impulsado por el grupo de artistas Computer Lars y el centro tecnológico MindFuture, quienes buscan que el partido se mantenga en el tiempo y se expanda globalmente. También proponen la creación de un nuevo Objetivo de Desarrollo Sostenible (ODS) enfocado en la relación entre humanos y robots (Vicente, 2023).

Otro ejemplo se dio en el Reino Unido, de cara a las elecciones de 2024, con el candidato de Smarter UK, AI Steve, un avatar de IA que representó al candidato legal Steve Endacott en la circunscripción de Brighton Pavilion. Su objetivo era buscar restaurar el interés de sectores apáticos en la política al permitir que los votantes influyeran directamente en las decisiones. De haber resultado electo, Endacott se ocuparía de representar

físicamente a AI Steve en el Parlamento británico, actuando según el voto mayoritario de los votantes. El proyecto fue objeto de varios cuestionamientos en torno a su ética y a su capacidad de cumplir eficazmente su cometido, y planteó además una serie de dudas legales. De todas formas, AI Steve se ubicó último en la circunscripción (un duro bastión urbano del Partido Verde) con tan solo 179 votos (Smith, 2024).

Conclusiones

Los cambios vividos en la forma de hacer campañas electorales en las últimas dos décadas son vertiginosos. Esta dinámica, aunada al deterioro de los partidos tradicionales y la emergencia de figuras disruptivas, carismáticas y populistas, ha significado un giro de 180 grados en materia de comunicación política.

La aparición de las redes sociales, la sofisticación de los métodos para recopilar, analizar y cruzar datos de manera masiva, perfilar a la audiencia (o el electorado) y diseñar mensajes específicos nos han llevado a otro juego, uno que no puede ser entendido a partir de las categorías antiguas.

Sin embargo, esta nueva realidad trae aparejados nuevos desafíos, incluso mayores que los de antaño. Si en las últimas décadas hemos padecido de una fatiga democrática, caracterizada por la desafección política, la crisis de representación y el descenso en la adhesión democrática de los ciudadanos, la irrupción del *bigdata* y la IA representa un reto no solo para los partidos y el electorado, sino para las instituciones, la integridad del espacio comunicacional y finalmente el sistema democrático.

Por otro lado, las nuevas tecnologías también han permitido conocer mejor al electorado, lo han complejizado y proble-

matizado. Esto puede ser una oportunidad para que los partidos tomen en cuenta las necesidades y preocupaciones de los electores, y que estas finalmente tengan impacto e influencia en sus programas y propuestas.

Por lo pronto, al parecer seguimos descubriendo los efectos de la inteligencia artificial en las campañas electorales y sería precipitado sacar conclusiones definitivas. Su uso comporta desafíos muy grandes y plantea debates éticos fundamentales, como por ejemplo si las personas son realmente libres de acceder a información plural, amplia y crítica, o, por el contrario, se encuentran cada vez más cercadas dentro de burbujas informativas específicamente diseñadas para moldear sus elecciones. ♦

Referencias bibliográficas

- BBC MUNDO. (2018, 21 de marzo). 5 claves para entender el escándalo de Cambridge Analytica que hizo que Facebook perdiera us\$37.000 millones en un día. <https://www.bbc.com/mundo/noticias-43472797>
- CEBRIÁN BELTRÁN, S. (2024). De la talla única al traje a medida: el *microtargeting* político para influir en las elecciones. Ponencia en el *XXI Congreso de la Asociación de Constitucionalistas de España*, mesa «Garantías constitucionales de elecciones libres», Valladolid. <https://www.acoes.es/wp-content/uploads/2024/02/CEBRIAN-BELTRAN-S.De-la-talla-unica-al-traje-a-medida-el-microtargeting-politico-para-influir-en-las-elecciones.pdf>
- DILLON, A. (2024, 26 de febrero). India confronts Google over Gemini AI tool's 'fascist Modi' responses. *The Guardian*. <https://www.theguardian.com/world/2024/feb/26/india-con>



La lucha anticorrupción desde el nuevo paradigma tecnológico

DENISSE RODRÍGUEZ-OLIVARI

La inteligencia artificial se ha convertido en una de las herramientas más novedosas en la lucha contra la corrupción. Esta nueva tecnología cuenta con innumerables bondades sin precedentes para identificar irregularidades, automatizar las comprobaciones de cumplimiento y mejorar la transparencia en las organizaciones públicas y privadas. Analizamos las oportunidades y retos de la aplicación de las tecnologías basadas en IA para detectar, prevenir y combatir la corrupción alrededor del mundo, con énfasis en América Latina.

Introducción

La lucha contra la corrupción continúa siendo una de las problemáticas más prevalentes en América Latina. De acuerdo al índice de control de corrupción elaborado por Americas Society/Council of the Americas (AS/COA) y Control Risks, un 70% de los expertos consideran que la corrupción continúa siendo un desafío enorme en sus países, solo superado por la inseguridad ciudadana y la situación económica pospandemia. Asimismo, según el índice de percepción de la corrupción de la ONG Transparencia Internacional, la mayoría de los países de la región empeoraron o se estancaron en materia anticorrupción. De hecho, solo Guyana y República Dominicana registran avances significativos y Venezuela cae a un mínimo histórico en el *ranking* global.

En medio de este escenario desolador, el uso de la inteligencia artificial (IA) ofrece soluciones novedosas a este problema tan complejo. Por un lado, están el análisis de datos y la detección de anomalías, la predicción de riesgos de corrupción y la automatización de procesos. Entendida como «una serie de sistemas que exhiben un comportamiento que analiza su entorno y toma acciones, con cierto grado de autonomía, para lograr ob-

jetivos específicos» (Grupo de Expertos de Alto Nivel sobre IA de la Comisión Europea, 2019), también es imperativo mencionar los riesgos de su uso poco ético. Por ejemplo, se reportan la erosión de los datos personales y vigilancia indiscriminada, la discriminación algorítmica y una implementación indebida en campañas electorales.

Sin embargo, las oportunidades parecen superar ampliamente a los riesgos. La proliferación de la IA como una tecnología anticorrupción —también conocida como AI-ACT por sus siglas en inglés— ha despertado el interés de los funcionarios públicos, los activistas y periodistas de investigación, así como los académicos expertos en la materia. La AI-ACT constituye una serie de ensamblajes sociotécnicos que permiten analizar grandes volúmenes de información, reducir la discrecionalidad de los funcionarios públicos y mediar entre los ciudadanos y los gobiernos (Mattoni, 2024).

Su potencial predictivo y preventivo aporta agilidad y eficiencia a la lucha contra la corrupción en América Latina, especialmente en Brasil, país líder en materia de IA para anticorrupción. Odilla (2023) sistematiza más de treinta iniciativas impulsadas por funcionarios públicos (*top down*) y organizaciones civiles (*bottom up*) que

promueven el monitoreo, identificación, reporte y predicción de irregularidades y riesgos de corrupción. Rosie, Alice y Monica son solo algunas de las iniciativas a la vanguardia de la IA-ACT en América Latina. Por ejemplo, Alice (Análisis de Licitaciones, Contratos y Convocatorias Públicas) y Monica (Monitoreo Integrado para el Control de Adquisiciones) escanean los procesos de adquisición para identificar irregularidades, y el Servicio de Impuestos Internos de Brasil utiliza la IA para detectar fraudes aduaneros, mientras que algunas agencias estatales están implementando la IA para esfuerzos anticorrupción más amplios. Sin embargo, hay una lenta adaptabilidad a nuevos tipos de irregularidades con una auditoría limitada a cargo de expertos. Aún queda por afinar la generación de datos sesgados, una implementación responsable y el perfeccionamiento de algoritmos fallidos.

Oportunidades y beneficios de la IA en la lucha contra la corrupción

La inteligencia artificial generativa y los grandes modelos de lenguaje ofrecen en el ámbito gubernamental un sinfín de oportunidades para innovar en la lucha contra la corrupción. A continuación, se mencionan algunas de sus aplicaciones.

Análisis de datos y reconocimiento de patrones

La IA permite procesar grandes volúmenes de información para detectar tempranamente posibles escenarios de actividad corrupta a través de inconsistencias o duplicidad de información en transacciones financieras, licitaciones, contrataciones o subvenciones. La IA-ACT puede ayudar a predecir áreas o sectores con mayor riesgo de fraude en el futuro basándose en datos

« La mayoría de los países de la región empeoraron o se estancaron en materia anticorrupción. Solo Guyana y República Dominicana registran avances significativos y Venezuela cae a un mínimo histórico en el ranking global. »

históricos y tendencias actuales. Esto es particularmente útil para identificar conflictos de interés o evaluar riesgos de corrupción, a fin de detectar anomalías e indicios de corrupción.

Un caso particular es el de la gobernanza proactiva en España a través del sistema Saler (Sistema de Alerta Rápida) de la Inspección General de Servicios de la Generalitat Valenciana. Este tiene como objetivo principal anticipar riesgos o debilidades que puedan perjudicar a la administración pública por inercia o malas prácticas. Saler se vale de la enorme cantidad de información digitalizada que maneja la Generalitat Valenciana, además de bases de datos de registradores, notarios, entidades de propiedad intelectual, entre otras, para analizar cualquier procedimiento administrativo de interés para oficiales de *compliance*. Estos riesgos se asocian a seguridad de información, licitaciones, comités de selección, colusión, verificaciones, gobernanza, ética, cumplimiento de la ley, recursos humanos, entre otros.

Mediante este sistema, por ejemplo, la seguridad social se apoya en algoritmos para monitorear individuos en situación de paro y, con ello, detectar incidencias de

fraude entre los beneficiarios de subvenciones. Esta herramienta predictiva se evalúa el estado de salud de las personas y predice la probabilidad de que estén en condiciones de volver al trabajo. Cabe señalar que debe apoyarse de un análisis más minucioso para evitar el desbordamiento de falsos positivos en desmedro de ciudadanos que realmente necesitan la ayuda estatal.

Del mismo modo, es posible analizar gastos de gobiernos locales o ministerios e identificar posibles sobrecostos, deficiencias o malas prácticas. Entre los países miembros de la OCDE, las compras públicas representan el 13% de su producto interno bruto, mientras que para América Latina dicho porcentaje alcanza el 8% (Pérez, 2021). Pese a que es una de las áreas más vulnerables a la corrupción, los recursos para el control y el monitoreo del gasto público suelen ser escasos. Se presentan a continuación dos casos de estudio al respecto.

VigIA

Desarrollado por el *Tic Tank* de la Universidad del Rosario —un *think tank* con enfoque en tecnologías de la información— junto a la Corporación Andina de Fomento (CAF) para la Veeduría Distrital de Bogotá, este ejemplo de uso de IA contribuye a la supervisión de contratos de Alcaldía de Bogotá con alto riesgo de corrupción e ineficiencias, aprovechando los datos proporcionados por el Sistema Electrónico para la Contratación Pública de Colombia (Secop) a través de la plataforma de la Agencia Nacional de Contratación Pública (Colombia Compra Eficiente). A través de modelos de aprendizaje de máquinas (*machine learning*) se predice el riesgo de incurrir en corrupción en cada contrato. Mediante esta tecnología

» La IA facilita la automatización de tareas rutinarias, lo que permite reducir el error humano, el tiempo requerido para identificar riesgos de corrupción [...] [y] mejorar la eficiencia de los procesos. «

se agiliza el trabajo de auditoría en la detección de ineficiencias e irregularidades que alimenta puntajes de riesgo con que las veedurías distritales determinan qué contratos tienen la mayor probabilidad de ser irregulares o ineficientes.

Serenata de amor: IA para el control social de la administración pública

Este proyecto, concebido en Brasil en 2016 por el científico de datos Irio Musskopf, el sociólogo Eduardo Cuducos y el empresario Felipe Cabral a raíz del estallido de uno de los escándalos más sonados de corrupción en el país, conocido como *Mensalao* —*gran mensualidad*, en castellano—, debido a los pagos indebidos al interior de la Cámara baja del Parlamento brasileño. Mediante el uso de *machine learning*, con datos del gobierno, se detectan señales de alerta en el gasto público. Estas incidencias son publicadas en X a través de un bot llamado Rosie, al que se suman La Denunciante (*whistleblower*, en inglés), Jarbas (que permite la visualización de datos) y Toolbox. La etapa inicial registró 629 denuncias que involucraban gastos de 216 diputados federales de un total de 513.



Optimización de procedimientos

La IA facilita la automatización de tareas rutinarias, lo cual permite reducir el error humano y el tiempo requerido para identificar riesgos de corrupción. Sirve principalmente para mejorar la eficiencia de los procesos y, por consecuencia, los esfuerzos anticorrupción. Por ejemplo, Aarvik (2019) cita el caso del gobierno de Kenya en colaboración con el grupo de investigación de IBM a fin de reducir los incentivos de recurrir a sobornos para realizar trámites. Recordemos que buena parte de las transacciones corruptas tratan de engrasar las ruedas, es decir, agilizar trámites que no deberían tomar tanto tiempo o recursos en primer lugar. Al hacer los procedimientos más eficientes, el país mejoró su posición en el *ranking* Doing Business del Banco Mundial que evalúa el ambiente regulatorio para hacer negocios en 189 países. Kenya pasó del

puesto 136 al 56¹. Es imperativo contar con un alto grado de digitalización en los gobiernos como paso previo a esta transformación.

También existen oportunidades para acelerar canales de denuncia, otro de los pilares de la anticorrupción. Utilizar la IA para diseñar e implementar un algoritmo de *priorización* y recepción de denuncias tiene como objetivo jerarquizarlas reduciendo los costos asociados a atender cada una de ellas por los funcionarios. Un estudio realizado por Pierri y Lafuente (2022), del Banco Interamericano de Desarrollo, aporta evidencia sobre el procesamiento de denuncias ciudadanas en el programa Nuevos Talentos en Control Gubernamental, de la Contraloría General de la República (CGR), en Perú.

¹ Después de irregularidades en los datos de los informes Doing Business de 2018 y 2020, la administración del Banco Mundial suspendió la elaboración de la edición siguiente e inició una serie de revisiones y auditorías del informe y su metodología.



Con una muestra de cinco mil incidentes, donde un 40% no requieren la atención de la CGR, mediante el uso de dos algoritmos de *priorización y admisión* se aumentó la tasa de éxito en el manejo de denuncias en un 36%, mientras que la tasa de éxito de las advertencias mejoró en un 27%. De acuerdo con estos resultados preliminares, el programa parece ser una iniciativa efectiva para mejorar los procesos internos de la CGR y contribuir a la lucha contra la corrupción en Perú. Mejorar el proceso de denuncia permite que la IA recomiende una serie de pasos o acciones basándose en casos anteriores, optimizando así los recursos y los tiempos de investigación.

Riesgos y desafíos del uso de la IA en la lucha contra la corrupción

Pese al creciente interés de la sociedad en las inmensas oportunidades del uso de la IA

y el consenso generalizado sobre el impacto positivo de esta en materia anticorrupción (Colonelli et al., 2020), existe también una serie de peligros directos e indirectos que deben considerarse. Figuras públicas del mundo de la tecnología, academia, funcionarios y periodistas firmaron una carta sobre los riesgos y estrategias de mitigación como una prioridad global, al mismo nivel que el manejo pospandemia o la guerra nuclear. Algunos de estos se mencionan brevemente a continuación.

Manipulación de los sistemas por actores corruptos

Podría existir una corrupción de IA donde el uso de sistemas de IA por funcionarios se realiza para obtener beneficios privados (Köbis et al., 2022). Siguiendo la lógica de la definición más popular de corrupción, los servidores públicos potencialmente tendrían acceso a una serie de ensambla-

jes tecnológicos que les permitiría cometer actos ilícitos y abusar de su discrecionalidad. Al tratarse de una nueva tecnología, la opacidad en el diseño, manipulación e implementación de esta puede obstaculizar la comprensión del proceso de toma de decisiones y, por consecuencia, generar desconfianza entre los usuarios. Y no estamos hablando de uso conscientemente corrupto de la IA, sino también aprovechamiento de las vulnerabilidades de los sistemas.

Riesgos de vigilancia masiva y violación de derechos civiles

Aunque la corrupción sigue siendo una preocupación primordial para los latinoamericanos, el principal lastre es la inseguridad. Aunque América Latina concentra el 9% de la población mundial, también registra la tercera parte de la tasa de homicidios a nivel global. Dentro de la región, México destaca con más de treinta mil asesinatos anuales en medio de una pelea por el control del territorio por una docena de carteles.

No es sorpresa que durante las últimas elecciones el ex candidato presidencial Marcelo Ebrard presentase un plan de seguridad llamado Angel (Avanzadas Normas de Geolocalización y Seguridad) que incluía cámaras con dispositivos de reconocimiento facial a fin de crear un ecosistema basado en IA con todas las bases de datos de México. Suponía la implementación de tecnologías de vigilancia masiva y biométrica en el espacio público para generar sistemas de vigilancia predictiva, iniciativas de ley para instalar cámaras de vigilancia con tecnología de reconocimiento, geolocalización de vehículos, identificación morfológica de delincuentes por la forma de caminar, uso de drones

» Si bien la IA tiene el potencial de ser una herramienta poderosa en la lucha contra la corrupción, es esencial abordar sus implicaciones éticas a fin de garantizar un uso responsable y transparente. «

y cámaras inteligentes corporales para elementos de la Guardia Nacional mexicana.

Sin embargo, numerosas investigaciones demuestran que las tecnologías de vigilancia masiva con reconocimiento facial son altamente susceptibles de identificar erróneamente a las personas, lo que ha provocado múltiples casos de detenciones arbitrarias de personas inocentes identificadas por equivocación como perpetradoras de delitos, especialmente si no son blancas.

Sobredependencia de la tecnología

Por otro lado, una dependencia excesiva de la tecnología puede hacer de una institución o un gobierno un blanco fácil de ciberataques y fallas técnicas. Es necesario tener en cuenta experiencias como Albania, que se convertiría en el primer país en desplegar el uso de la IA en su proceso de adhesión a la Unión Europea. El gobierno albanés utilizará ChatGPT, el modelo de LLM más popular del mundo con un millón de usuarios en su primera semana, para traducir al shqip miles de páginas de medidas y disposiciones legales de la UE, a fin de integrarlas en la legislación albanesa vigente. Esto ocurre luego de un acuerdo con OpenAI, cuya directora tecnológica eje-

cutiva Mira Murati es de origen albanés. Sin duda, se trata de una manera novedosa que ahorrará dinero y tiempo al aparato estatal. Aunque no hay que olvidar las implicaciones éticas sobre el vacío legal existente en materia de privacidad, transparencia y dependencia excesiva de la tecnología en dicho país.

Sesgos algorítmicos en la identificación

Ya que esta tecnología se alimenta de información preexistente, no es más que un reflejo fidedigno de lo que ocurre en la realidad (que también es susceptible de sesgos conscientes e inconscientes). A esto se le llama *sesgo algorítmico*, cuya repercusión en el uso de IA anticorrupción es generar falsos positivos mediante el refuerzo de desigualdades y discriminaciones. Por ejemplo, si al analizar la incidencia de la corrupción en el pasado existió una concentración de personas con características particulares de grupo étnico, edad, lugar de origen u ocupación, un algoritmo destinado a detectar futuros casos de corrupción podría correlacionar erróneamente a personas con dichas características pero que no han cometido actos ilícitos.

Estudios en Estados Unidos muestran que mientras más oscuro sea el color de piel, menos acertado será el resultado. Esto se debe a que los repositorios existentes cuentan, en su mayoría, con una sobrerrepresentación de hombres blancos. Esto es particularmente problemático en el uso policial. El uso de datos de entrenamiento del algoritmo con información sesgada no hace más que generar fallos en la identificación y, finalmente, decisiones de tratamiento incorrectas. Por ello, es vital utilizar la tecnología a favor de la eficiencia y la cobertura, pero sin prescindir de la pericia y experiencia de agentes encargados de filtrar y analizar las bases de datos que alimentan los sistemas de IA.

Reflexiones finales en torno al potencial de la IA en la lucha contra la corrupción y el futuro de la gobernanza pública en la era digital

Si bien la IA tiene el potencial de ser una herramienta poderosa en la lucha contra la corrupción, es esencial abordar sus implicaciones éticas a fin de garantizar un uso responsable y transparente. Saber aprovechar las bondades de la IA para analizar, predecir y automatizar datos, ya sea desde las agencias anticorrupción o desde la sociedad civil, debe ser acompañado de mecanismos que promuevan la transparencia, la rendición de cuentas y la mitigación de sesgos que aseguran una buena gobernanza de IA.

Cabe señalar los diferentes retos y oportunidades para las iniciativas en materia de AI-ACT que vengan desde las agencias anticorrupción (desde arriba) o desde la sociedad civil (desde abajo). Las primeras podrían perpetuar las existentes asimetrías de poder de las entidades estatales y generar consecuencias no anticipadas en la lucha contra la corrupción, mientras que las segundas se enfrentan a barreras de acceso a datos abiertos que dependen en gran medida del nivel de digitalización de los gobiernos. Sin embargo, apoyadas de las redes sociales y la vigilancia ciudadana, ofrecen ventajas para la difusión y actualización en tiempo real de incidencias y riesgos de corrupción.

Por otro lado, si bien las iniciativas IA anticorrupción agilizan procesos, estos no pueden prescindir de la acción humana. La opacidad en el diseño e implementación de algoritmo sigue siendo un desafío para generar confianza en los sistemas, sobre todo en América Latina, donde la mayoría de la gente tiene bajos niveles de confianza interpersonal y en las instituciones. De acuerdo con un estudio del Banco Interamericano de Desarrollo,

nueve de cada diez latinoamericanos desconfían de su prójimo. Esto se suma a la falta de independencia y transparencia del Poder Judicial en la mayoría de los países de la región, así como a vacíos y limitaciones de la legislación vigente y el grado de discrecionalidad de las élites políticas y económicas. En un contexto de este tipo, es complejo implementar mecanismos anticorrupción que se valgan o no de la IA, que sean eficaces y cuenten con el respaldo ciudadano necesario.

Más allá de innovaciones, los organismos sancionadores también deben operar con sumo cuidado nuevos sistemas, a fin de evitar agravar la desigualdad y generar una persecución selectiva basada en sesgos algorítmicos. Por ejemplo, los funcionarios públicos que provienen de comunidades vulnerables o de minorías podrían ser desproporcionadamente investigados o sancionados, aunque sus acciones no sean más corruptas que las de otros funcionarios. Esto no solo crea una injusticia individual, sino que también refuerza los estereotipos sociales y agrava la percepción de parcialidad en la lucha contra la corrupción.

Si las personas perciben que el sistema solo persigue a ciertos grupos, en lugar de tratar a todos los funcionarios por igual, puede generarse una desconfianza hacia las instituciones anticorrupción. La falta de transparencia sobre cómo se toma la decisión de investigar a alguien también puede contribuir a esta desconfianza. Se registran casos donde la IA es utilizada por regímenes autoritarios y de manera discrecional como la persecución de activistas antirracismo en Miami y Nueva York, grupos feministas en Estados Unidos, grupos feministas en la Ciudad de México, activistas prodemocracia en Hong Kong, periodistas y opositores políticos en Egipto,

« Si las personas perciben que el sistema solo persigue a ciertos grupos, puede generarse una desconfianza hacia las instituciones anticorrupción. »

to, hasta el monitoreo sistemático, perfilamiento y persecución de minorías étnicas como los uigures en Xinjiang.

Una de las estrategias de mitigación de riesgo es la transparencia algorítmica para explicar cómo funcionan los algoritmos y qué tipo de decisión toman, de forma que puedan ser auditables y monitoreables y así eliminar la perpetuación de sesgos. Del mismo modo opera el aumento en número y variedad de las fuentes de datos, para entrenar los algoritmos mientras se protege la privacidad de la información personal de los usuarios.

El camino a seguir es generar más confiabilidad y precisión en las predicciones de riesgos, sin reemplazar la pericia y experiencia de auditores. Esto representa una gran tarea en América Latina, región con una gran brecha de infraestructura digital, donde solo el 57% de los ciudadanos tienen acceso a un servicio activo de internet móvil. La diferencia entre América del Sur y Centroamérica es casi del doble (77 % en la primera, y 37% en la segunda), destacando Brasil (77%) y Haití tan solo 6%. Cerrar las brechas entre y dentro de cada país, considerando las desigualdades entre las zonas rurales y urbanas, es vital para asegurar una implementación y despliegue justo de la arquitectura digital de sistema de IA para la lucha contra la co-

rrupción, así como sistemas neutrales que reduzcan los sesgos y las desigualdades producto de estos.

Los gobiernos, las empresas y la sociedad civil deben comprometerse en aprovechar el poder transformativo de la IA mediante una vigilancia constante, nuevo marco regulatorio y compromiso ético. Recordemos que corromper a un individuo puede tener un impacto en un número limitado de personas. En cambio, corromper un algoritmo podría afectar a miles de personas en tiempo récord. ♦

Bibliografía

- AARVIK, P. (2019). *Artificial Intelligence – a promising anti-corruption tool in development settings?* U4 Anti-Corruption Resource Centre. <https://www.u4.no/publications/artificial-intelligence-a-promising-anti-corruption-tool-in-development-settings>
- ADAM, I., Y FAZEKAS, M. (2021). Are emerging technologies helping win the fight against corruption? A review of the state of evidence. *Information Economics and Policy*, 57, p. 100950. DOI: 10.1016/j.infoecopol.2021.100950.
- COLONNELLI, E., GALLEGO, J. A., Y PREM, M. (2020, 26 de diciembre). *What Predicts Corruption?* https://www.researchgate.net/publication/347995419_What_Predicts_Corruption
- DÁVILA PÉREZ, J. (2021). Impacto y beneficios de las reformas en los sistemas de contratación pública en América Latina y el Caribe. *Red Interamericana de Compras Gubernamentales*. <https://ricg.org/es/publicaciones/impacto-y-beneficios-de-las-reformas-en-los-sistemas-de-contratacion-publica-en-america-latina-y-el-caribe/>

KÖBIS, N. (2023). Bribes for Bias: Can AI be corrupted? *Transparency International Blog*. <https://www.transparency.org/en/blog/bribes-for-bias-can-ai-be-corrupted>

KÖBIS, N., STARKE, CH., Y EDWARD-GILL, J. (2022). *The corruption risks of artificial intelligence*. Transparency International Working Paper. <https://knowledgehub.transparency.org/assets/uploads/kproducts/The-Corruption-Risks-of-Artificial-Intelligence.pdf>

ODILLA, F. (2023). Bots against corruption: Exploring the benefits and limitations of ai-based anti-corruption technology. *Crime Law Soc Change*, 80(4), 1-44. DOI: 10.1007/s10611-023-10091-0.

PIERRI, G., Y LAFUENTE, M. (2022). *Human Talent Management and Corruption Control: The Effect of the New Talents in Government Control Program on the Detection of Corruption in Peru*. IADB Discussion Paper, IDB-DP-952. <https://publications.iadb.org/en/human-talent-management-and-corruption-control-effect-new-talents-government-control-program>



Denisse Rodríguez-Olivari

Doctora en Ciencia Política (Humboldt-Universität, Berlín). Máster en Desarrollo Internacional (Universidad de Manchester). Licenciada en Ciencia Política y Gobierno (Pontificia Universidad Católica del Perú). Research Associate de la University of Glasgow, Adam Smith Business School. Experta en anticorrupción e integridad.

X: @DenisseROlivari

IG: @denisser

LI: denisserodriguezolivari



Inteligencia artificial y seguridad pública en las democracias latinoamericanas

GUSTAVO FLORES-MACÍAS Y BÁRBARA HERNÁNDEZ

La inteligencia artificial revoluciona diversos ámbitos de las políticas públicas. Su aplicación en el combate al crimen permea el discurso político contemporáneo y aparece en las propuestas de seguridad pública de los candidatos presidenciales. Garantizar que la IA se utilice de forma responsable y ética es fundamental para maximizar sus beneficios y mitigar los riesgos que plantea a las libertades civiles en las democracias latinoamericanas.

Los recientes avances en inteligencia artificial (IA) están revolucionando diversos ámbitos de políticas públicas. La seguridad pública no es la excepción, ya que la IA y el aprendizaje automático se aplican a la vigilancia policial y al cumplimiento de la ley en general. Varias aplicaciones de la IA para la seguridad pública, definida como «la aplicación de algoritmos a vastas cantidades de datos para asistir a la labor policial» (Joh, 2018) han recibido considerable atención en los Estados Unidos y otras democracias industrializadas. Entre los usos más destacados se encuentran los sistemas de reconocimiento facial (que capturan los rasgos físicos únicos de las personas y clasifican millones de posibilidades para establecer la identidad de un sospechoso), los sistemas de lectura de matrículas (que pueden capturar, identificar y encontrar su coincidencia con vehículos y sus propietarios para determinar su paradero), la predicción de la posible ubicación de futuros delitos y hasta la prevención de delitos de fraude en línea y pruebas forenses de ADN.

Algunas de estas aplicaciones de la IA siguen siendo controvertidas debido a consideraciones éticas y de privacidad. Aunque su adopción muestra un gran po-

tencial para hacer frente a la delincuencia, la IA también presenta desafíos importantes para las democracias liberales donde la protección de las libertades civiles es primordial.

A pesar de estas preocupaciones, muchos países latinoamericanos han recurrido a la rápida adopción de tecnologías de IA para combatir a la delincuencia. Los partidarios de estas tecnologías señalan la velocidad, la capacidad de analizar inmensas cantidades de datos que serían inasequibles para el ser humano, así como la reducción de errores y sesgos humanos como importantes beneficios que justifican la adopción generalizada de la IA para la seguridad pública.

Teniendo en cuenta que las tecnologías de lucha contra el crimen basadas en IA siguen siendo poco estudiadas en América Latina en comparación con los Estados Unidos, este texto presenta una introducción a su adopción en las democracias de la región. En las siguientes secciones, nos centramos en ejemplos de IA utilizada para mejorar la seguridad pública en América Latina. A través de estos ejemplos, ilustramos sus propósitos previstos, así como las oportunidades y los desafíos que representan.

La promesa de la IA para la seguridad pública en América Latina

El uso de la IA para combatir la delincuencia en América Latina es prometedor. Con apenas un 8% de la población mundial, la región concentra alrededor del 30% de los homicidios a escala global. Además, si bien hay una variación considerable en términos de tasas de delitos violentos entre países (desde Chile con una tasa de alrededor de 4,5 homicidios por cada 100.000 personas, hasta Ecuador con una tasa de alrededor de 44 por cada 100.000), los promedios nacionales pueden ocultar tasas considerablemente más altas al interior de los países (Manjarrés y Newton, 2024).

En este contexto, la tecnología basada en IA puede ayudar a las instituciones encargadas de la aplicación de la ley a realizar varias tareas de manera más eficiente y eficaz. En particular, la IA puede ayudar a recopilar y analizar cantidades masivas de datos sobre delitos y predecir posibles focos de actividad delictiva, lo cual permite desplegar recursos de manera más eficaz a las dependencias encargadas de hacer cumplir la ley.

La IA también puede permitir a las policías de la región realizar un seguimiento en tiempo real para mantener la seguridad pública. Las herramientas de IA pueden ayudar a monitorear cámaras de vigilancia, redes sociales u otras fuentes abiertas de información para detectar actividades sospechosas y alertar a las autoridades. De manera similar, la IA se puede aprovechar para el reconocimiento facial y la biometría con el fin de identificar a sospechosos o personas desaparecidas de manera más eficiente, particularmente en entornos complicados o multitudinarios.

En el ámbito forense, las herramientas de IA pueden desempeñar un papel im-

«**Muchos países latinoamericanos han recurrido a la rápida adopción de tecnologías de IA para combatir a la delincuencia.**»

portante en la gestión y el análisis de pruebas. Son valiosas para organizar y filtrar grandes cantidades de pruebas digitales y así mejorar la eficiencia de las investigaciones y los procedimientos judiciales. De este modo, la IA tiene el potencial de ayudar a abordar los sesgos humanos y orientar la toma de decisiones en función de datos objetivos para lograr resultados más transparentes y justos.

Ejemplos de uso de IA en América Latina

Debido a la creciente preocupación por la delincuencia en la región, los gobiernos latinoamericanos han buscado mejorar la seguridad pública mediante la incorporación de avances tecnológicos basados en IA. Si bien esta adopción ha sido producto de ensayo y error, varios ejemplos de su uso son alentadores. A medida que aumenta el número de casos y los gobiernos y ciudadanos se acostumbren a la tecnología, es probable que su adopción se acelere cada vez más.

A todos los niveles, los gobiernos de la región han incorporado la IA para garantizar la seguridad pública. En Colombia, la Policía Nacional publicó una estrategia nacional para brindar recursos a los gobiernos locales para comprar drones no tripulados y cámaras de vigilancia para prevenir y detectar delitos (Ministerio del

Interior..., 2020). Para conectar las extensas bases de datos sobre criminalidad con sus diversos puntos de referencia para tomar decisiones, el gobierno contrató un software externo de Amazon Web Services para organizar y almacenar datos de varias fuentes («IA al servicio de la seguridad pública...», 2023). Además de analizar los resultados rápidamente, el software también emplea una serie de herramientas, incluida la aplicación XCrime de Nuvu, para agregar y analizar información, así como predecir posibles delitos (Nuvu, 2023).

En Chile, la Asociación de Aseguradores desarrolló un software de IA para identificar y denunciar vehículos robados (Municipio de Teno, 2024). El software convierte los teléfonos móviles en lectores de matrículas que los departamentos de policía pueden utilizar fácilmente. Identifica automáticamente vehículos robados escaneando las matrículas y comparándolas con una base de datos en segundos. El programa se implementó inicialmente en 60 municipios y se amplió a 345 en 2022 (Asociación de Aseguradores de Chile, 2022).

También se han realizado esfuerzos a nivel estatal y local. En México, la Oficina de las Naciones Unidas contra la Droga y el Delito (Unodc) se asoció con el organismo de estadísticas del país, el Inegi, para desarrollar un programa en el estado de México que aprovecha las redes neuronales profundas para identificar palabras y frases clave en llamadas realizadas al número 911 con sospecha de violencia doméstica o delitos contra la mujer (Unodc México, 2023). A nivel municipal, los gobiernos locales también han realizado esfuerzos para incorporar IA, especialmente en programas de vigilancia asistida (Vázquez Cruz, 2023).

El gobierno de Benito Juárez, uno de los 16 municipios de la Ciudad de México,

» los gobiernos latinoamericanos han buscado mejorar la seguridad pública mediante la incorporación de avances tecnológicos basados en IA. Si bien esta adopción ha sido producto de ensayo y error, varios ejemplos de su uso son alentadores. «

adoptó una estrategia de seguridad pública en 2018 para abordar la delincuencia a través de herramientas de vigilancia basadas en IA (Tenorio, 2018). Este gobierno aumentó el número de unidades policiales y creó una amplia red de vigilancia compuesta por cámaras del Gobierno y cámaras privadas entregadas a ciudadanos particulares como parte de kits de seguridad pública («App “Blindar Benito Juárez”...», 2021). Las imágenes se transmiten a un centro de control y comando (C2) que emplea el reconocimiento facial y de matrículas para identificar y rastrear a sospechosos («Benito Juárez presenta...», 2021). El programa también aprovecha la participación ciudadana al brindar acceso a Blindar BJ, una aplicación donde los ciudadanos pueden denunciar delitos y rastrear la ubicación de las unidades de policía que responden a su denuncia. En 2021, el programa se expandió a la vecina alcaldía Álvaro Obregón.

En 2018, el gobierno municipal de Tlajomulco de Zúñiga, en el estado de Jalisco, también buscó mejorar la seguridad

pública al integrar tecnología basada en IA en los sistemas de vigilancia, incluido el monitoreo de áreas públicas con cámaras de videovigilancia así como la puesta a disposición a los ciudadanos de información sobre la incidencia delictiva («Tlajomulco presenta...», 2018; Gobierno de Jalisco, 2018). El esfuerzo involucra 564 cámaras de reconocimiento facial y lectores de matrículas para monitorear el comportamiento, identificar patrones y reportar actividades sospechosas.

En Brasil, el gobierno de la ciudad de São Paulo inauguró en 2022 el uso de un sistema de reconocimiento facial basado en inteligencia artificial para monitorear la línea 3 del sistema de metro a través de 14.000 cámaras en 18 estaciones (Mascellino, 2022). La ciudad también está expandiendo su vigilancia basada en biometría más allá del metro, a través del programa Smart Sampa que vinculará unas 20.000 cámaras con un centro de monitoreo y con bases de datos de otras agencias para tareas que van desde la identificación de sujetos hasta la localización de ambulancias. Aunque el proyecto ha recibido críticas por posibles sesgos, en abril de 2024 el alcalde de la ciudad anunció la inauguración del centro de monitoreo y espera completar la instalación de las 20.000 cámaras de vigilancia para fines de 2024 (Freitas, 2024; Mari, 2023).

En Uruguay, cuando Montevideo comenzó a registrar mayores tasas de delincuencia, el departamento de policía de la ciudad adoptó en 2016 una herramienta de vigilancia predictiva llamada PredPol. Aprovechando datos históricos sobre la delincuencia, el software generó mapas de predicciones delictivas para informar los esfuerzos de la policía y así prevenir el delito o responder rápidamente a las amenazas emergentes. PredPol fue imple-

» La IA puede utilizarse para monitorear la actividad legal de los ciudadanos, como sucede en regímenes autoritarios como China.«

mentada en Montevideo con base en la herramienta utilizada por primera vez por el Departamento de Policía de Los Ángeles en California para generar predicciones delictivas basadas en el tipo de delito, ubicación, fecha y hora. Con un costo de USD 140 millones anuales y con un algoritmo propio al que los gobiernos no tienen acceso, este software ayudó a agilizar el despliegue policial y abordar el delito de manera preventiva. Sin embargo, el Gobierno finalizó el programa después de que no encontró diferencias entre las zonas que utilizaban el software predictivo y las que no (Gaudín, 2017).

Desafíos para la adopción de la IA para la seguridad pública

Si bien estos ejemplos muestran el interés de los gobiernos de la región por usar esta herramienta para combatir los crecientes niveles de delincuencia, su potencial y su uso plantean desafíos importantes. Frente a la presión pública para obtener resultados, las autoridades han actuado rápidamente para adoptar dichas tecnologías, pero a menudo lo hacen sin una legislación adecuada que asegure la protección de la privacidad y las libertades civiles, que promueva procesos de contratación transparentes y que garantice la sustentabilidad de la IA. En esta sec-

ción se analizan algunos de los principales desafíos que enfrentan los gobiernos en la adopción de tecnología de IA para la seguridad pública, así como las medidas adoptadas para abordar estas deficiencias.

Vacíos regulatorios

La regulación de la IA es crucial para prevenir su uso indebido y anticipar consecuencias no deseadas. Si bien la mayoría de los países latinoamericanos carecen de legislación sobre IA en general, y para fines de seguridad pública en particular, algunos están en proceso de generar marcos regulatorios (por ejemplo, Argentina y Brasil) basados en regímenes multilaterales como la Ley de IA de la Unión Europea y la Organización de Cooperación y Desarrollo Económico (OCDE), o con el apoyo del Banco Interamericano de Desarrollo. Sin embargo, mientras los reguladores lidian con resguardar la protección de los derechos sin comprometer el crecimiento económico, no se ha alcanzado consenso sobre cómo abordar la regulación de la IA (EGA, 2024; OECD.AI Policy Observatory, 2022).

Por ejemplo, en Brasil existen dos propuestas legislativas regulatorias: la primera impone menos restricciones a la IA al apuntar a un sistema descentralizado y restringir la intervención gubernamental, mientras que la segunda impone limitaciones al uso de la IA y establece un sistema de categorización de sistemas de IA con base en su riesgo (EGA, 2024). Si bien la legislación sobre IA es menos común, algunos países como Brasil, Chile y Colombia han publicado estrategias nacionales de IA que proveen lineamientos para impulsar su adopción (OECD.AI Policy Observatory, 2021). Además, varios países latinoamericanos firmaron recientemente la Declaración de Santiago, que fomenta una participación

más activa en las deliberaciones sobre IA (Cumbre Ministerial..., 2023).

Privacidad y libertades civiles

Aunque la IA puede ser una herramienta eficaz para combatir la delincuencia, también representa importantes riesgos para los derechos humanos. Por ejemplo, un uso común es la vigilancia de espacios públicos para identificar patrones de comportamiento potencialmente dañinos y prevenir delitos. Si bien esto puede ser beneficioso, este monitoreo constante puede convertir a cualquier persona en un sospechoso potencial, incluso cuando no se ha cometido ningún delito, ya que todas las acciones cotidianas pueden ser objeto de escrutinio en busca de un crimen.

En este contexto, una preocupación importante es el posible uso de la IA para socavar las protecciones que se brindan a los ciudadanos en el marco de la democracia liberal. En particular, la IA puede utilizarse para monitorear la actividad legal de los ciudadanos, como sucede en regímenes autoritarios como China. El reconocimiento facial, los lectores de matrículas o los transmisores de ubicación de teléfonos celulares pueden utilizarse para rastrear a personas, incluso cuando no han cometido un delito.

Algunos ejemplos de la región son ilustrativos. En Argentina, la policía registró por error las variables equivocadas para identificar a un sospechoso y, en su lugar, encarceló a un hombre inocente durante seis días (Naundorf, 2023). En el Programa Blindar Benito Juárez de la Ciudad de México, algunos usuarios notaron la falta de un aviso de términos de privacidad al descargar la aplicación del gobierno, lo que generó inquietudes sobre la privacidad de los usuarios («Sin trans-

parencia...», 2021). En Ecuador, informes señalan que la agencia de inteligencia del gobierno recurre al sistema de vigilancia del delito ECU-911 importado de China para espiar a periodistas y miembros de la oposición con el fin de obtener ventajas políticas (Kessel, 2019). Estos incidentes muestran cómo la tecnología de inteligencia artificial puede socavar las libertades civiles si los gobiernos la utilizan de manera incorrecta.

Adquisición y corrupción

Los procesos de compra gubernamentales en América Latina suelen ser opacos, y la adquisición de tecnología basada en IA no es la excepción. En contraste con la velocidad con que los gobiernos de la región han adoptado la IA para la seguridad pública, la transparencia en los procesos de adquisición ha sido lenta (Access Now, 2021). Un informe de AccessNow concluyó que algunas empresas extranjeras que suministraban servicios de IA para seguridad pública en América Latina negaban vender herramientas de vigilancia, reformulaban su propósito o desviaban la responsabilidad hacia los usuarios finales (Access Now, 2023).

En el caso de Blindar Benito Juárez y Álvaro Obregón, en la Ciudad de México, la falta de transparencia en la asignación de fondos ha generado dudas. Como referencia, el gobierno invirtió 385 millones de pesos mexicanos (unos 19,6 millones de dólares) en un lapso de tres años, pero ha enfrentado acusaciones de que el valor de los equipos está inflado (Díaz, 2023). Si bien una investigación estimó que cada cámara cuesta alrededor de MX\$ 2.700 (USD 159), el gobierno las presupuestó en MX\$ 35.000 (USD 2.000) (Cosme, 2021).

Adopción y mantenimiento de tecnología

Además de requerir marcos regulatorios más robustos, los países latinoamericanos también tienden a carecer de infraestructura suficiente para garantizar que las tecnologías funcionen correctamente. En Bogotá, el gobierno de la ciudad adoptó un sistema basado en IA para identificar y predecir el crimen. Con una inversión considerable de 11 millones de dólares, el sistema generó preocupaciones sobre los derechos humanos y la falta de policías para actuar en función de la información generada (Valencia Gómez, 2021), pero en particular fue criticado porque se estimó que el 22% de sus cámaras no funcionaba correctamente («Sistema de videovigilancia ...», 2022). Además, las diferencias en el software utilizado en las distintas plataformas de IA limitaron el intercambio de información y la eficiencia, como fue el caso de los software para el sistema de tránsito rápido de autobuses públicos Transmilenio y para el centro de comando central C4 de las cámaras de vigilancia de la ciudad.

De manera similar, en la Ciudad de México, el programa Blindar Benito Juárez enfrentó importantes desafíos de implementación. En particular, los vecinos se quejaron de que las cámaras no grababan, lo que dificultaba la presentación de pruebas cuando se producían delitos (Gobierno de la Ciudad de México, 2023). Desafíos como estos podrían influir en la percepción de la utilidad de la IA si no son atendidos.

Toma de decisiones de caja negra

Los algoritmos empleados en la IA no suelen ser conocidos por las agencias gubernamentales. En la mayoría de los casos, las autoridades tendrán poco conocimiento sobre cómo se ponderan los datos de entrada o hasta qué punto pueda haber sesgos.



« la IA para combatir el crimen está permeando el discurso político contemporáneo, e incluso está presente en las propuestas de seguridad pública de los candidatos presidenciales. »

En el caso de Montevideo, los altos costos (USD 140 millones por año), la dependencia de un conjunto fijo de datos de entrenamiento y los sesgos en el reporte de datos históricos sobre delitos fueron preocupaciones importantes (Gaudín, 2017; Lum, 2016). Dado que el algoritmo se alimentaba de datos históricos policiales, críticos señalaron que el modelo podría dirigir la atención a áreas ya altamente vigiladas, creando un ciclo de

retroalimentación sesgado. Además, dada la naturaleza patentada del software, las autoridades no tuvieron acceso a los algoritmos. De manera similar, con el proyecto Smart Sampa de Brasil, las organizaciones de derechos humanos han destacado su potencial para encarcelar con mayor frecuencia a las minorías y comunidades marginadas.

La IA en el ojo público

A medida que aumenta el uso de tecnología de IA para la aplicación de la ley en América Latina, también aumenta su presencia en el ámbito público. En particular, la IA para combatir el crimen está permeando el discurso político contemporáneo, e incluso está presente en las propuestas de seguridad pública de los candidatos presidenciales. Por ejemplo, durante las elecciones presidenciales de Panamá en 2024, varios candidatos propusieron aprovechar la IA para prevenir

el crimen. Durante su campaña, el presidente José Raúl Mulino propuso capacitación adicional y la adquisición de nueva tecnología de IA para prevenir el crimen («Plan de Gobierno...», 2024), y el expresidente Martín Torrijos abogó durante su campaña por el uso de sistemas de vigilancia y monitoreo con inteligencia artificial junto con el aumento de las unidades de la policía («A otro nivel...», 2024).

De manera similar, los candidatos presidenciales en México también han hecho campaña sobre el uso de la IA para frenar la delincuencia, desde la instalación del mayor número posible de cámaras de vigilancia hasta la expansión de los programas de seguridad pública que han aprovechado dichas tecnologías. Como lo expresó la candidata de centroderecha en las elecciones presidenciales de 2024, «vamos con todo en el uso de tecnología y de inteligencia» (Rangel, 2024). Aunque estos esfuerzos parecen prometedores, es preocupante que los candidatos pongan más énfasis en la implementación o expansión de programas sin considerar también las políticas necesarias para regular estas tecnologías y garantizar su uso responsable.

A pesar de la controvertida historia de los países latinoamericanos en materia de vigilancia gubernamental de la sociedad civil, no parece haber una fuerte oposición a la tecnología. De hecho, encuestas recientes muestran que los ciudadanos a menudo tienen una perspectiva positiva hacia la IA. En 2023, Ipsos, una firma global de investigación de mercados, encuestó a más de 22.000 personas en 31 países, incluidos cuatro de América Latina: México, Colombia, Chile y Argentina, y les preguntó sobre las percepciones públicas de la IA (Ipsos, 2023).

A nivel mundial, alrededor del 67% de los encuestados estaban muy o algo de

acuerdo en que entendían bien la IA. En comparación, los promedios en América Latina fueron generalmente más altos (México 75%, Colombia 73%, Chile 70% y Argentina 67%). De manera similar, mientras que el 54% de los encuestados en todo el mundo estaban muy o algo de acuerdo en que los productos y servicios de IA brindan más beneficios que desventajas, la proporción de encuestados que compartían esta opinión fue mayor en América Latina: 73% en México, 65% en Colombia, 59% en Chile y 57% en Argentina. Cabe destacar que las actitudes hacia la IA entre los países latinoamericanos fueron generalmente más favorables que las de sus contrapartes europeas.

Conclusión

La adopción de tecnologías basadas en la IA para la seguridad pública en las democracias de América Latina ofrece importantes posibilidades de abordar las altas tasas de delincuencia de la región y mejorar la eficiencia de la aplicación de la ley. Desde la vigilancia predictiva hasta el reconocimiento facial y la vigilancia en tiempo real, las herramientas de IA tienen el potencial de transformar los esfuerzos de lucha contra el delito al mejorar el análisis de datos, el despliegue de recursos y las investigaciones forenses en una región donde el tráfico de drogas y el crimen organizado en general están en aumento. Sin embargo, la rápida integración de estas tecnologías también presenta desafíos sustanciales para las democracias liberales, incluidas lagunas regulatorias, preocupaciones sobre la privacidad y cuestiones de transparencia en las contrataciones públicas. El riesgo de que se socaven las libertades civiles, junto con la falta de infraestructura y mantenimiento adecuados, subraya la necesidad de un enfoque

más cauteloso y regulado para la implementación de la IA.

A medida que la IA adquiere mayor protagonismo en el discurso político y en las estrategias de seguridad pública, los gobiernos deben priorizar el establecimiento de marcos jurídicos sólidos que equilibren la innovación con la protección de los derechos humanos. Garantizar que la IA se utilice de forma responsable y ética será fundamental para maximizar sus beneficios y mitigar los riesgos que plantea a las libertades civiles en las democracias latinoamericanas. ♦

Referencias bibliográficas

- A otro nivel. Plan de gobierno 2024-2029 Martín Torrijos. (2024). <https://static.tvn-2.com/tvn/public/content/file/original/2024/0324/22/martin-torrijos-pa-nama-a-otro-nivel-plan-de-gobierno-2024-2029-pdf.pdf>
- ACCESS NOW. (2021, 10 de agosto). *Made Abroad, Deployed at Home*. <https://www.accessnow.org/wp-content/uploads/2021/08/Surveillance-Tech-Latam-Report.pdf>
- ACCESS NOW. (2023, abril). Remote biometric surveillance in Latin America. Are companies respecting human rights? <https://www.accessnow.org/wp-content/uploads/2023/04/ENG-Analysis-Remote-biometric-surveillance-LATAM.pdf>
- App “Blindar Benito Juárez” te permite ver cámaras de seguridad; así funciona. (2021, 8 de abril). *El Heraldo de México*. <https://heraldodemexico.com.mx/nacional/2021/4/8/app-blindar-benito-juarez-te-permite-ver-camaras-de-seguridad-asi-funciona-281458.html>
- ASOCIACIÓN DE ASEGURADORES DE CHILE. (2022, 22 de junio). Asociación de Aseguradores firma convenio con municipios para detectar y denunciar autos robados. <https://portal.aach.cl/noticias/asociacion-de-aseguradores-firma-convenio-con-municipios-para-detectar-y-denunciar-autos-robados/>
- Benito Juárez presenta su centro digital de vigilancia. (2021, 7 de abril). *El Universal*. <https://www.eluniversal.com.mx/metropoli/benito-juarez-presenta-su-centro-digital-de-vigilancia/>
- COSME, M. (2021, 17 de noviembre). Lía Limón solicita 5 mil 784 mdp de presupuesto 2022 para Álvaro Obregón. *El Sol de México*. <https://www.elsoldemexico.com.mx/metropoli/cdmx/lia-limon-solicita-5-mil-784-mdp-de-presupuesto-2022-para-alvaro-obregon-7488301.html>
- CUMBRE MINISTERIAL Y DE ALTAS AUTORIDADES DE AMÉRICA LATINA Y EL CARIBE. (2023, 24 de octubre). Declaración de Santiago. https://minciencia.gob.cl/uploads/filer_public/40/2a/402a35a0-1222-4dab-b090-5c81bbf34237/declaracion_de_santiago.pdf
- DÍAZ, O. (2023, 13 de noviembre). Miguel Hidalgo, Álvaro Obregón y Azcapotzalco piden aumento de presupuesto para el 2024. *El Universal*. <https://www.eluniversal.com.mx/metropoli/miguel-hidalgo-y-alvaro-obregon-piden-aumento-de-presupuesto-para-el-2024/>
- EGA. (2024). *Artificial Intelligence. Latin America's Regulatory and Policy Environment*. https://www.edelmanglobaladvisory.com/sites/g/files/aatuss676/files/2024-03/EGA%20LATAM_ai%20Policy%20in%20Latin%20America_4Mar2024_0.pdf
- FREITAS, H. (2024, 4 de julio). Smart Sampa: Prefeitura inicia programa de câmeras com reconhecimento facial em São Paulo. *O Globo*. <https://oglobo.globo.com/brasil/sao-paulo/noticia/2024/07/04/>

- smart-sampa-prefeitura-inicia-programa-de-cameras-com-reconhecimientofacial-em-sao-paulo.ghtml
- GAUDÍN, A. (2017, 5 de mayo). Uruguay Tries Preventative Policing with a High-tech Twist. <https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=15513&context=notisur>
- GOBIERNO DE JALISCO. (2021). Plan Municipal de Desarrollo y Gobernanza. Tlajomulco de Zúñiga. <https://seplan.app.jalisco.gob.mx/biblioteca/archivo/descargarArchivo/3174>
- GOBIERNO DE LA CIUDAD DE MÉXICO. (2023, 28 de marzo). *Síntesis informativa. Alcaldía*. <https://www.jefaturadegobierno.cdmx.gob.mx/storage/app/media/Sintesis%20Informativa/28032023m-alcaldias-institutos-oi-congreso.pdf>
- IA al servicio de la seguridad pública: el éxito de la Policía Nacional de Colombia prediciendo crímenes en AWS. (2023). AWS. <https://aws.amazon.com/es/solutions/case-studies/policia-nacional-colombia/>
- IPSOS. (2023, julio). Global Views on A.I. 2023. How people across the world feel about artificial intelligence and expect it will impact their life. https://www.ipsos.com/sites/default/files/ct/news/documents/2023-07/Ipsos%20Global%20AI%202023%20Report-WEB_0.pdf
- JOH, E. (2018). Artificial Intelligence and Policing: First Questions. *Seattle University Law Review*, vol. 41, 1139-1144. <https://digitalcommons.law.seattleu.edu/cgi/viewcontent.cgi?article=2550&context=sulr&fbclid=IwAR2nc1q2Qf2OHTkOb5A5cYhvACuLuGlbPudtTxdEuIUPMec5akh3EKFFEXU>
- KESSEL, J. (2019, 26 de abril). In a Secret Bunker in the Andes, a Wall That Was Really a Window. *The New York Times*. <https://www.nytimes.com/2019/04/26/reader-center/ecuador-china-surveillance-spying.html>
- LUM, K. (2016, 10 de octubre). Predictive Policing Reinforces Police Bias. *hrdag*. <https://hrdag.org/2016/10/10/predictive-policing-reinforces-police-bias/>
- MANJARRÉS, J., Y NEWTON, CH. (2024, 21 de febrero). InSight Crime's 2023 Homicide Round-Up. *InSight Crime*. <https://insightcrime.org/news/insight-crime-2023-homicide-round-up/>
- MARI, A. (2023, 13 de julio). Facial recognition surveillance in São Paulo could worsen racism. <https://www.aljazeera.com/economy/2023/7/13/facial-recognition-surveillance-in-sao-paulo-could-worsen-racism>
- MASCELLINO, A. (2022, 9 de diciembre). Brazil deploys iss facial recognition to secure São Paulo metro. *biometricupdate.com*. <https://www.biometricupdate.com/202212/brazil-deploys-iss-facial-recognition-to-secure-sao-paulo-metro>
- Ministerio del Interior, Ministerio de Defensa Nacional, Consejería Presidencial para la Seguridad Nacional. (2020). Política Marco de Convivencia y Seguridad Ciudadana 2019-2022. *Dirección de Antinarcóticos*. https://oas.policia.gov.co/sites/default/files/ciena-politica_1.pdf
- Municipio de Teno. (2024). Municipio de Teno instala lectores de patentes para detectar autos robados. <https://www.teno.cl/webteno/index.php/2024/05/03/municipio-de-teno-instala-lectores-de-patentes-para-detectar-autos-robados/>
- NAUNDORF, K. (2023, 15 de septiembre). Un escándalo en Buenos Aires revela los peligros del reconocimiento facial. *Wired*. <https://es.wired.com/articulos/escandalo-en-buenos-aires-revela-los-peligros-del-reconocimiento-facial>
- NUVU. (2023). Policía Nacional de Colombia. <https://www.nuvu.cc/caso-de-exito/policianacionaldecolombia>
- OECD.AI Policy Observatory. (2021). Natio-

- nal AI policies & strategies. <https://oecd.ai/en/dashboards/overview>
- OECD.AI Policy Observatory. (2022, 6 de septiembre). Brazilian ai Strategy. <https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faiipo.oecd.org%2F2021-data-policyInitiatives-27104>
- Plan de Gobierno del candidato a la presidencia José Raúl Mulino. (2024). <https://static.tvn-2.com/tvn/public/content/file/original/2024/0405/15/plan-de-gobierno-jose-raul-mulino-pdf.pdf>
- RANGEL, L. (2024, 7 de marzo). Tecnología en propuestas de seguridad de candidatas: uso de datos biométricos y videovigilancia pone en riesgo derechos humanos. *El Sabueso*. <https://animalpolitico.com/verificacion-de-hechos/te-explico/propuestas-seguridad-tecnologia-videovigilancia>
- Sin transparencia el funcionamiento de las aplicaciones de seguridad de la BJ. (2021, 25 de febrero). *DDM Benito Juárez*. <https://ddmbj.mx/sin-transparencia-el-funcionamiento-de-las-aplicaciones-de-seguridad-de-la-bj>
- Sistema de videovigilancia de la Ciudad ¿en cuidados intensivos? (2022). Concejo de Bogotá. <https://concejodebogota.gov.co/sistema-de-videovigilancia-de-la-ciudad-en-cuidados-intensivos/cbogota/2023-10-24/163021.php>
- TENORIO, G. (2018, 6 de mayo). Santiago Taboada hace público su Plan de Gobierno 2018-2021 para la alcaldía de Benito Juárez. *Periódico Leo*. <https://periodicoleo.com/ccdmx/pan/2018/05/06/santiago-taboada-hace-publico-su-plan-de-gobierno-2018-2021-para-la-alcaldia-de-benito-juarez/>
- Tlajomulco presenta modelo de vídeo vigilância. (2018). *Tlajo*. <https://tlajomulco.gob.mx/noticias/tlajomulco-presenta-modelo-de-video-vigilancia>
- Unodc México. (2023, 30 de enero). Inteligencia artificial para detectar y prevenir la violencia contra las mujeres. <https://www.unodc.org/lpomex/es/noticias/enero-2023/inteligencia-artificial-para-detectar-y-prevenir-la-violencia-contra-las-mujeres.html>
- VALENCIA GÓMEZ, M. (2021, 26 de enero). El reto de anticipar delitos con tecnología en Bogotá. *El Espectador*. <https://www.elespectador.com/bogota/el-reto-de-anticipar-delitos-con-tecnologia-en-bogota-article/>
- VÁSQUEZ CRUZ, E. (2023, 27 de septiembre). Inteligencia Artificial para mejorar la seguridad pública, una tendencia mundial. *Alcaldes de México*. <https://www.alcaldesdemexico.com/de-puno-y-letra/inteligencia-artificial-para-mejorar-la-seguridad-publica-una-tendencia-mundial/>



Gustavo Flores-Macías

Profesor de Gobierno Comparado y Políticas Públicas en la Universidad de Cornell, Estados Unidos. Investigador afiliado del Cornell Tech Policy Institute.

X: @Gustavo_F_M

LI: gustavo-flores-macias



Bárbara Hernández

Licenciada en Relaciones Internacionales y Psicología por la Universidad de Stanford. Investigadora independiente.

A portrait of Lars Zimmermann, a man with short, light-colored hair, wearing a dark sweater. He is looking directly at the camera with a neutral expression. The background is a blurred indoor setting with blue and white lighting, possibly a hallway or office.

Lars Zimmermann:
«La tecnología no nos
eximirá de las tareas
de la democracia»

La IA actual sigue requiriendo de personas que trabajen con ella y la controlen y conduzcan. ¿Cuáles son las tareas específicas de la política en términos de buen gobierno con este nuevo avance tecnológico? Conversamos sobre el efecto de la IA en la transformación del Estado con un experto dedicado a la modernización e innovación de la administración pública.

Lars Zimmermann es cofundador y miembro del consejo de GovTech Campus Deutschland e.V., el primer espacio mundial de innovación, desarrollo y aprendizaje para la modernización del gobierno y la administración. El gobierno alemán impulsa el Campus como plataforma conjunta con el objetivo de desarrollar innovaciones y tecnologías digitales para la administración federal, estatal y local junto con la escena tecnológica. Busca poner este insumo a disposición para su reutilización.

Antes de fundar el GovTech Campus, Zimmermann se desempeñó como consultor en tecnología y transformación y fue fundador y portavoz de la junta de la Stiftung neue Verantwortung. Lleva muchos años implicado en los ámbitos de la modernización del Estado y la reforma administrativa, y ha desarrollado un gran número de medidas y proyectos en este ámbito. Desde comienzos de 2024 se desempeña como investigador asociado de la Fundación Konrad Adenauer.

—Con la llegada de la IA, ¿estamos a punto de dar un salto similar al de la llegada de la máquina de vapor?

—Sinceramente, aún no lo sabemos. Siempre soy un poco cauteloso a la hora de proclamar una gran revolución. Sin embar-

go, sospecho que la IA tendrá un impacto igualmente significativo en la sociedad. Estamos a punto de dar un salto en el desarrollo que habría sido impensable hace 15 años. Hace cuatro años, ninguno de nosotros hablaba de grandes modelos lingüísticos como ChatGPT y menos aún trabajaba con ellos.

Se ha producido un gran avance disruptivo en la IA, que por primera vez ha llegado a la población en general, a distintos niveles. Si extrapolamos este avance, tenemos que llegar a la conclusión de que la IA, como la máquina de vapor u otras grandes tecnologías, podría suponer un gran salto en el desarrollo. Por supuesto que esto implica riesgos, pero, sobre todo, también hay numerosas oportunidades. Y, en mi opinión, esta evolución llega en el momento justo.

—¿Por qué llega en el momento oportuno?

—Por ejemplo, en todo el mundo, especialmente en los países industrializados, tenemos el problema de que aumenta la brecha demográfica. Si desarrollamos y utilizamos adecuadamente la inteligencia artificial, podremos colmar estas lagunas en muchas profesiones. Un segundo ejemplo es la mejora de las capacidades informáticas. Con la inteligencia artificial y el fortalecimiento de las infraestructuras informáticas, actualmente podemos procesar, gestionar y

contextualizar datos, capacidades de las que no disponíamos hace diez años. Estos avances aportan grandes beneficios, por ejemplo, en la investigación sanitaria y en cuestiones de seguridad.

Por eso creo que, si bien estamos en una época de grandes retos, también tenemos grandes oportunidades gracias a la IA. Estamos experimentando grandes saltos tecnológicos que pueden ayudarnos a vencer enfermedades, desarrollar medicamentos y compensar la escasez de personal.

—**¿De qué periodo de tiempo estamos hablando ahora?**

—Es una buena pregunta que nadie puede responder hoy con certeza. Tal vez podamos formularlo de la siguiente manera: los saltos innovadores de la IA a corto plazo suelen sobreestimarse considerablemente, mientras que se subestiman mucho los que se producen a mediano y largo plazo. Si miramos hacia atrás, hace cuatro años nadie sabía lo que era un gran modelo lingüístico. Estoy convencido de que veremos enormes saltos en el desarrollo en los próximos 15 a 20 años.

—**¿Y a corto plazo?**

—A corto plazo, por ejemplo, los grandes modelos lingüísticos pueden suponer un aumento de la eficiencia. Estos modelos ya son muy potentes y capaces de asumir muchas tareas. Sin embargo, no pueden, por ejemplo, sustituir asesores en un ministerio federal. La IA actual sigue requiriendo de personas que trabajen con ella y la controlen

y conduzcan. Evidentemente, esto podría cambiar cuando la tecnología y las capacidades informáticas estén tan avanzadas que la IA pueda asumir determinadas tareas sin ayuda humana.

Naturalmente, estos cambios también nos afectarán a nosotros. La IA podría hacerse cargo de actividades que actualmente son realizadas por humanos, haciendo que los puestos de trabajo correspondientes sean superfluos. Pese a ello, no creo que la IA provoque un desempleo masivo. Por lo contrario, veo a la IA como una forma de subsanar las carencias causadas por la escasez de mano de obra. No se trata de reducir recursos, sino de llenar lagunas.

—**A menudo se percibe a la IA como una amenaza, y se la asocia frecuentemente con la pérdida de puestos de trabajo, las noticias falsas y los deepfakes. Entonces, ¿cuáles son las oportunidades?**

—Por lo pronto, la IA puede mejorar significativamente los procesos administrativos. La IA proporciona al personal administrativo una herramienta que le permite cumplir mejor sus tareas y más rápidamente. Un buen ejemplo de ello son los grandes modelos lingüísticos, que permiten redactar y sintetizar textos con rapidez.

Supongamos que los asesores y asesoras ministeriales tengan que crear plantillas. A menudo, un ministro o una ministra solicita que un texto largo se reduzca a una página concisa. En el pasado, esto solía llevar mucho tiempo. Con los grandes modelos lingüísticos actuales, este proceso puede automatizarse fácilmente. Lo que antes llevaba una hora, ahora puede resolverse en unos segundos. Esto ahorra muchísimo tiempo y aumenta la eficacia.

Otro ejemplo es esta misma entrevista, que realizamos en alemán pero que después se imprimirá en español. En el pasado, la traducción habría costado mucho dinero y

» La IA, como la máquina de vapor u otras grandes tecnologías, podría suponer un gran salto en el desarrollo. «

» La IA actual sigue requiriendo de personas que trabajen con ella y la controlen y conduzcan. «

tiempo. Hoy en día, gracias a grandes modelos lingüísticos como ChatGPT, se pueden traducir textos de forma rápida y eficaz a cualquier idioma. Esto simplifica y acelera considerablemente todo el proceso.

—Sin embargo, no podemos sustituir al traductor que empleamos actualmente. Necesitamos a alguien que revise las soluciones que se han desarrollado con los modelos lingüísticos.

—No me parece mal que un humano siga controlando las traducciones. Sin embargo, es probable que la tecnología de la traducción siga desarrollándose de tal manera que el control humano deje de ser necesario en algún momento. La tasa de error podría entonces ser comparable a la de un traductor simultáneo, que tampoco está libre de errores. Es probable que esta evolución se produzca con relativa rapidez. Sin embargo, no creo que las personas traductoras se queden sin trabajo. Cuando se trata de analizar e interpretar el lenguaje y evaluar matices que indiquen confianza en uno mismo o inseguridad en la voz, la IA aún no está preparada. Estas habilidades requieren percepción humana, así que pasará un tiempo antes de que la IA esté al mismo nivel en este ámbito. Pero es muy probable que lo consiga en algún momento.

—¿No significa esto que se sustituyen sobre todo tareas rutinarias de calificación media, con lo que se suprimen muchos puestos de trabajo, mientras que persiste la escasez de especialistas altamente calificados?

—Aún no estoy seguro de eso. Creo que los puestos de nivel medio tienden a correr más riesgo que los de nivel muy bajo. Cuanto más exigente sea un puesto de trabajo, es decir, de calificación media o superior, más probable es que sea sustituido por la IA. Esto afecta también a muchas instancias de control en la administración, que actualmente llevan a cabo especialistas más calificados, pero que podrían automatizarse en el futuro. Por ejemplo, los puestos de trabajo desempeñados por académicos en el sector de los seguros también podrían verse amenazados por la IA. Así que no se trata solo de trabajos sencillos, sino también de aquellos que requieren un mayor nivel de calificación. En el siguiente paso, la IA no hará obsoleto ningún trabajo, pero cambiará los perfiles de requisitos tradicionales de esos empleos. Quien no esté de acuerdo con esto, lo tendrá difícil.

—Nos gustaría saber su opinión sobre cómo afectará el desarrollo de la IA a Latinoamérica.

—En última instancia, el desarrollo de la IA sigue abierto. No se trata tanto de lo buenos que sean los distintos países en el desarrollo de la IA como de las capacidades de las empresas en este ámbito. En la actualidad, solo hay un puñado de empresas que impulsan la IA a escala mundial y la mayoría de ellas proceden de Estados Unidos.

El impacto de la IA en los distintos sistemas sociales aún no está claro. Las oportuni-

» En el siguiente paso, la IA no hará obsoleto ningún trabajo, pero cambiará los perfiles de requisitos tradicionales de esos empleos. «





dades y los riesgos están tan presentes en los países industrializados como en otras partes del mundo. Tomemos África como ejemplo: muchos países africanos lograron mayores avances que Alemania en cuanto a sistemas de pago y apoyo digital a las microempresas. Esto muestra que incluso las economías menos desarrolladas pueden beneficiarse de los saltos tecnológicos.

Si economías como las de América del Sur o América Central siguen desarrollándose, existe un riesgo real de que las innovaciones apoyadas en la IA conduzcan también a una mayor eficiencia laboral y, por tanto, a una reducción de puestos de trabajo. Esto significa que los gobiernos de dichas economías deben examinar de cerca el impacto de la IA para evitar quedarse atrás y para aprovechar en forma óptima las oportunidades que ofrece la tecnología.

—**Las mayores empresas del sector de la IA proceden actualmente de Estados Unidos. El llamado efecto red se aplica especialmente en este sector, donde el líder del mercado se beneficia enormemente y deja poco espacio a los competidores.**

—Existe esta famosa frase: «En el campo de la inteligencia artificial ya no hay un tercer puesto». O eres el líder del mercado mundial o eres el que le pisa los talones. Pero este riesgo se aplica a todos los países. En mi opinión, no se trata de una cuestión típica de la política de desarrollo.

Es importante entender que estos retos no pueden ser abordados por un solo país. Especialmente en regiones como Sudamérica, sería muy importante que los países trabajaran juntos a nivel supranacional y suprarregional para abordar la cuestión de la IA. Sin embargo, sigue siendo necesaria una buena gobernanza. Los países mal gobernados seguirán teniendo dificultades para utilizar las innovaciones tecnológicas en el futuro. Una gobernanza de alta cali-

«**Los gobiernos [de economías como las de América del Sur o América Central] deben examinar de cerca el impacto de la ia para evitar quedarse atrás.**»

dad basada en los valores democráticos fundamentales será aún más decisiva frente a dinámicos avances tecnológicos. La IA no puede hacerse cargo de esta tarea de buen gobierno; esta labor debe seguir en manos de los ciudadanos y probablemente adquiriera mayor importancia.

—**¿Cuáles son las tareas específicas de la política en términos de *buen gobierno* con estos nuevos avances tecnológicos?**

—En primer lugar, creo que los políticos deben desarrollar una gran curiosidad por la IA. Los países que triunfan en tiempos de innovación suelen ser los que se muestran abiertos y curiosos ante las nuevas tecnologías. Es importante no tener miedo a la tecnología ni demonizarla. La apertura a lo nuevo es crucial.

En segundo lugar, es esencial buscar la cooperación. Ningún país, ni siquiera Estados Unidos, puede abordar la IA por sí solo y con la profundidad requerida. Incluso países grandes como Estados Unidos y China se ven confrontados a desafíos. Para los países más pequeños como Alemania y Francia es aún más importante trabajar juntos más allá de las fronteras. Por lo tanto, es mucho más importante un alto grado de voluntad para cooperar y trabajar juntos en estos temas.

El tercer paso también es muy importante: crear las condiciones marco que permitan la innovación.

El cuarto paso es la aplicación real de la IA. Es importante que los países y sus ciudadanos no demonicen la IA, sino que la utilicen activamente. Quien utiliza IA también puede opinar. Un ejemplo del pasado: los alemanes se convirtieron en líderes mundiales del sector automovilístico porque el mercado nacional era fuerte y los propios alemanes compraban masivamente coches. En el campo de la IA, todavía no hay ningún país que haya tomado la delantera de forma creíble en su aplicación, ya sea en la educación o en la salud. Por lo tanto, todas las regiones aún siguen teniendo puntos de partida similares. Tal vez los países que aún no están tan saturados y que tienen menos estructuras posean incluso una ventaja, porque pueden estar especialmente abiertos a nuevos desarrollos.

El quinto paso se refiere a la formación. Para todos los temas, incluida la IA, la educación es crucial. El desarrollo y la aplicación de la IA y otras tecnologías no tendrán éxito si no contamos con centros de formación y sistemas educativos de primera clase. Es esencial que contemos con buenas instituciones educativas capaces de preparar a las próximas generaciones para las nuevas tecnologías e integrarlas en los sistemas educativos.

—¿Cuáles serían las narrativas positivas que los partidos políticos podrían contar sobre la IA?

—Hay muchas historias positivas que pueden contarse sobre la IA. Por ejemplo, puede dar respuesta a nuestros retos demográficos haciendo más eficientes los servicios públicos. Podría ayudar a reducir la necesidad de personal, lo que a largo plazo reduciría los costos administrativos. También podría suponer una mejor gestión del dinero de los contribuyentes e incluso una reducción de los impuestos.

» El desarrollo y la aplicación de la ia y otras tecnologías no tendrán éxito si no contamos con centros de formación y sistemas educativos de primera clase.«

Otra ventaja es la ganancia en materia de innovación. La IA permite procesar grandes cantidades de datos y utilizar eficazmente los centros de procesamiento de datos. Esto puede dar lugar a investigaciones pioneras (disruptivas, revolucionarias) en el sector de la salud. Estoy convencido de que viviré para ver cómo el cáncer deja de ser una sentencia de muerte. Con el avance de la potencia de cálculo y el procesamiento de datos sanitarios podríamos desarrollar fármacos que derroten enfermedades que hace veinte o treinta años se consideraban incurables.

Tomemos el ejemplo de la guerra en Ucrania. Sin las capacidades de IA, Ucrania no podría utilizar sus drones de forma tan eficiente y eficaz para su defensa, a pesar de estar tan mal equipada en cuanto a disponibilidad de armas. Esta capacidad de IA en combinación con los drones es crucial para la defensa del país.

Estos ejemplos demuestran que la IA puede aportar grandes avances en diversos ámbitos políticos. Esencialmente, podemos distinguir dos grandes áreas: la IA lleva a una mayor eficiencia y menores costos, y fomenta la innovación y las soluciones en diversos ámbitos.

—¿Llevará la IA también a nuevos formatos de participación en una democracia representativa?

—Antes de responderle, quisiera decirle que está hablando con alguien muy

partidario de la forma tradicional de democracia. Estoy convencido de que la tecnología no nos liberará de las tareas esenciales de la democracia. ¿Qué quiero decir con esto? No creo que la IA nos haga automáticamente mejores demócratas o que nos convierta en votantes mejor informados. A mis ojos, la democracia es un sistema que funciona entre personas e instituciones. Las personas siempre serán los actores centrales de una democracia, tomando decisiones por mayoría de votos. Por eso soy conservador, en el mejor sentido de la palabra. No creo que los sistemas democráticos de IA vayan a salvar la democracia.

—En la campaña electoral estadounidense vimos que los votantes se están volviendo transparentes para los estrategas gracias al análisis de grandes cantidades de datos. ¿Qué impacto tiene esto en la democracia?

—Esto no es necesariamente problemático. Demócratas fuertes hacen fuertes a las democracias. Usted alude al hecho de que los partidos conocen mejor a sus votantes y pueden colocarles anuncios específicos. Pero esto es parte de un desarrollo que siempre existió. Si miramos atrás en la historia económica, con la invención de la imprenta tipográfica de repente se pudieron imprimir y distribuir en masa octavillas, y estoy seguro de que la primera octavilla política apareció no mucho después de la invención de la imprenta tipográfica.

El principio sigue siendo el mismo: las tecnologías nos ayudan a difundir información más rápidamente y en tiempo real. ¿Es un riesgo que los políticos sepan más sobre la población? No necesariamente. Se convierte en peligroso cuando los Estados lo saben todo y están autorizados a combinar esta información dispersa en un bloque de información global y a utilizarla

» Es decisivo cuestionarse las cosas. Por ejemplo, por qué ves anuncios de un coche en Instagram después de haber hablado de él con tus amigos. ¿Acaso el smartphone está escuchando? «

para dar forma a la política. También aquí podemos ver que, en última instancia, la buena gobernanza será decisiva.

Se podría culpar indirectamente de ello a la IA, porque sin ella el Estado no podría recopilar esta información en la misma medida. Pero hay que ser capaz de resistir a la innovación. La democracia ha sobrevivido a la imprenta y a la invención de la televisión, y también sobrevivirá a la IA.

—¿Qué pueden hacer las personas para prepararse?

—Por de pronto, es importante mantener la curiosidad y la mente abierta. Mientras uno se involucre activamente en los nuevos avances, estará más informado y seguro de sí mismo. El mayor peligro es volverse perezoso y dejar que los demás le indiquen constantemente qué tiene que pensar.

Es decisivo cuestionarse las cosas. Por ejemplo, por qué ves anuncios de un coche en Instagram después de haber hablado de él con tus amigos. ¿Acaso el smartphone está escuchando? ¿Puede reconocer esos fragmentos de frases? Este tipo de preguntas te ayudan a ser crítico con las nuevas tecnologías y a tomar tus propias decisiones.

Es importante desarrollar una actitud crítica y constructiva. No tiene sentido rechazar inmediatamente todo lo nuevo

ni adoptarlo acríticamente. Por el contrario, hay que intentar formarse una opinión equilibrada y analizar críticamente lo nuevo.

También es útil probar cosas. Quien entiende cómo funciona algo, puede decidir mejor si quiere usarlo o no. Alemania ha alcanzado el liderazgo en innovación en muchas áreas porque la gente estaba dispuesta a probar cosas nuevas. También en el campo de la IA, la gente debe probar y comprender diferentes aplicaciones.

Por ejemplo, hoy ya utilizamos la IA tal vez sin darnos cuenta. Mi banco utiliza un sencillo sistema de IA para reconocer movimientos sospechosos en mi cuenta. Este tipo de IA nos protege y facilita muchas tareas cotidianas.

En resumen, la apertura a nuevas ideas, el pensamiento crítico y la experiencia práctica son claves para encontrar el camino en la era de la IA y reconocer las oportunidades y los riesgos para uno mismo. De este modo, cada individuo puede contribuir a que la sociedad y el país avancen por este camino.✦

Traducción alemán-español: Manfred Steffen



Jonathan Neu

Representante adjunto del Programa Regional Partidos Políticos y Democracia en América Latina, de la Fundación

Konrad Adenauer, con sede en Montevideo, Uruguay. Estudió matemáticas e historia en las universidades de Leipzig y Salamanca. Se especializó en historia de las ideas.

X: @NeuJonathan



Manfred Steffen

Magíster en Ciencias Ambientales por la Universidad de la República de Uruguay. Ingeniero en Imprenta por la FHD

Stuttgart, Alemania. Redactor de Diálogo Político y coordinador de proyectos del Programa Regional Partidos Políticos y Democracia en América Latina, de la Fundación Konrad Adenauer.

X: @ManfredSteffenN



La libertad y la verdad secuestradas por la inteligencia artificial

MIGUEL PASTORINO

La inteligencia artificial es una nueva realidad con la que convivimos y que transforma también lo que somos, cómo nos pensamos y cómo vivimos, como un *nuevo medio ambiente*. Su impacto es transformador y radical, desde la educación hasta la medicina, desde la economía y la política hasta el trabajo y los vínculos entre las personas.

La llamada inteligencia artificial no es solamente un salto tecnológico, sino que ha generado un cambio antropológico, porque redefine la vida humana, las formas de pensar y vivir, de conocer, de aprender, de relacionarse, las formas del poder y los límites de los derechos, la concepción de la libertad y de la verdad. Representa una de las cuestiones filosóficas más relevantes de nuestro tiempo. Porque no es una simple herramienta, sino que comienza a fusionarse con nuestro ambiente y con quienes somos, transformando nuestra vida cotidiana y su impacto no deja ambiente sin transformar, desde la educación hasta la medicina, desde la economía y la política, hasta el trabajo y los vínculos entre las personas.

A la vez que hemos moldeado a las máquinas, las mismas máquinas nos han moldeado a nosotros. El mundo construido por el ser humano, el mundo de las máquinas o los sistemas sociotécnicos, nos ha rediseñado en parte afectando a nuestras habilidades y dando forma a nuestros propios valores y creencias (Savulescu y Lara, 2021).

La IA no es ni moral ni filosóficamente neutra, no es un instrumento, sino una nueva realidad con la que convivimos y que

transforma también lo que somos, cómo nos pensamos y cómo vivimos, un *nuevo medio ambiente*. Nos encontramos en una dinámica de interacción mutua, de coparticipación de operaciones.

El ambiente tecnológico tiene sus propias dinámicas y su propia lógica, no ha reemplazado el ambiente *natural*, sino que lo ha transformado, lo ha reconfigurado totalmente: «La tecnología ya no es un medio, sino que es el Medio, el ambiente en el que vivimos. Se trata de un Medio porque, debido a su autorregulación y sus automatismos, se configura como algo independiente y en continua evolución, capaz de rodearnos y de crear siempre nuevos contextos para el ser humano... Las nuevas tecnologías interactúan con el ambiente que las rodea, respondiendo a estímulos que llegan a ellas y que modifican sus comportamientos de manera independiente...». (Varela, 2022)

Que la tecnología abra nuevos espacios de acción implica necesariamente abrir una discusión sobre esas acciones, que nunca son neutras. Nos exige pensar en las consecuencias sociales, económicas y políticas, pero antes es necesario preguntarse por lo

« La IA no es ni moral ni filosóficamente neutra, no es un instrumento, sino una nueva realidad con la que convivimos y que transforma también lo que somos, cómo nos pensamos y cómo vivimos, un nuevo medio ambiente. »

que realmente es y su impacto sobre la condición humana.

¿De qué inteligencia hablamos?

La inteligencia humana es irreducible a las funciones que puedan compararse metafóricamente con las que le atribuimos a la IA. Confundir funciones con la singularidad humana es un reduccionismo bastante habitual. De hecho, nos tienen acostumbrados al uso prestado de nociones tomadas de las ciencias cognitivas para aplicarlo a los sistemas informáticos, sin las debidas aclaraciones conceptuales. Los términos que usamos forjan nuestras representaciones mentales: Se habla de *chips sinápticos*, *redes de neuronas artificiales*, procesadores neuronales, etc.

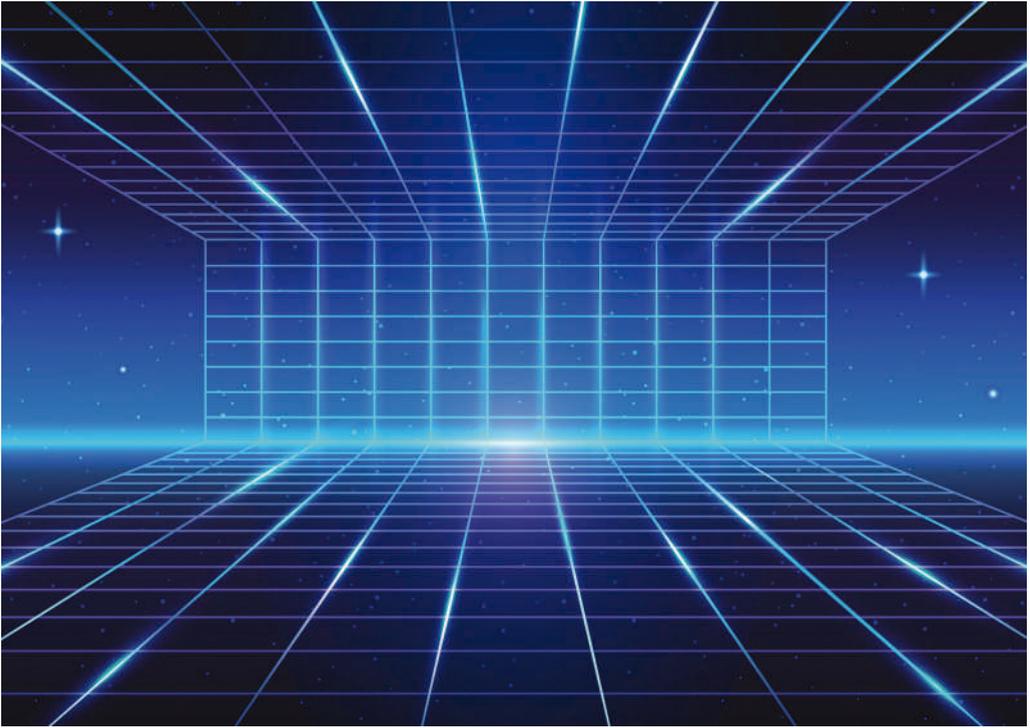
«El principio de una inteligencia computacional modelada sobre nuestra inteligencia es erróneo, porque una y otra no mantienen casi ninguna relación de similitud» (Sadin, 2020), salvo que se caiga en la tentación de reducir la inteligencia a algunas funciones y que se reduzca la realidad a códigos binarios excluyendo así una infinidad de dimensiones que nuestra subjetividad humana puede

experimentar y que escapan a una modelización matemática. «Estamos frente a una concepción trunca, restringida y sesgada de lo que supone el proceso de la inteligencia, que es indisoluble de su tensión con una aprehensión multisensorial y no sistematizable del medio ambiente» (Sadin, 2020).

El *big data* apareció como una forma de conocimiento absoluto, a partir del cual las cosas revelan sus correlaciones secretas, pero olvidamos preguntarnos por el sentido de las cosas, por el por qué, por la razón última de los acontecimientos y por el significado de la vida.

«Todo se vuelve calculable, predecible, controlable. Se anuncia toda una nueva era del saber. En realidad, se trata de una forma de saber bastante primitiva. La minería de datos (*data mining*) descubre correlaciones. Según la lógica de Hegel, la correlación representa la forma más baja del saber» (Han, 2021), porque con las correlaciones no sabemos por qué suceden las cosas, simplemente sabemos que suceden.

A pesar de los impresionantes avances con la inteligencia artificial generativa y nuevas transformaciones en ciencia y tecnología, lo cierto es que no hablamos de inteligencia en el sentido humano, porque aunque la IA pueda realizar por aprendizaje automático una serie de funciones que hacemos los humanos (cálculo, procedimientos matemáticos, selección de información, detección de patrones, reproducción de lo aprendido) en un tiempo y cantidad de información imposible para un ser humano, eso no significa que piense humanamente. La IA no tiene conciencia ni subjetividad, aunque pueda simular emociones e interactuar con humanos aprendiendo y reaccionando a la información que recibe. El problema es cuando reducimos inteligencia a cálculo y manejo de información. Las máquinas no producen sabiduría porque



« La IA no tiene conciencia ni subjetividad, aunque pueda simular emociones e interactuar con humanos aprendiendo y reaccionando a la información que recibe. »

no tienen subjetividad, no tienen autoconciencia, aunque puedan simularla y hacernos creer que sí e impresionarnos.

La IA no es artificial ni inteligente. Más bien existe de forma corpórea, como algo material, hecho de recursos naturales, combustible, mano de obra, infraestructuras, logística, historias y clasificaciones. Los sistemas de IA no son autónomos,

racionales ni capaces de discernir algo sin un entrenamiento extenso y computacionalmente intensivo, con enormes conjuntos de datos o reglas y recompensas predefinidas. De hecho, la IA como la conocemos depende por completo de un conjunto mucho más vasto de estructuras políticas y sociales. Y debido al capital que se necesita para construir IA a gran escala y a las maneras de ver que optimiza, los sistemas de IA son, al fin y al cabo, diseñados para servir a intereses dominantes ya existentes. (Crawford, 2022)

Sumisión al oráculo artificial

Con el pretexto de tomar *las mejores decisiones* en todas las esferas de la vida (financieras, transporte, salud, deporte, justicia, etc.), los asuntos humanos comienzan

a resolverse desde las alturas de la *superinteligencia* artificial, desde donde se tienen y procesan mayor cantidad de datos. Asistimos a una creciente consulta de oráculos artificiales que hacen las veces de gurús o directores espirituales que imponen rutinas diarias como poseedores de un conocimiento superior e incuestionable. Y así podemos ver cómo todo puede empezar por un nivel básico, de *coaching*, donde una aplicación opera sobre la vida emocional, sobre la nutrición o los vínculos, pautando cómo pensar y actuar, hasta niveles más prescriptivos donde la IA decidirá el futuro laboral o si se es digno de un préstamo bancario. Sadin refiere a un nivel más radical al que se ha llegado, un nivel coercitivo donde la IA terminará decidiendo gastos, recortes e incluso a impartir justicia.

La humanidad se está dotando a grandes pasos de un órgano de prescindencia de ella misma, de su derecho a decidir con plena conciencia y responsabilidad las elecciones que la involucran. Toma forma un estatuto antropológico y ontológico inédito que ve cómo la figura humana se somete a las ecuaciones de sus propios artefactos con el objetivo prioritario de responder a intereses privados y de instaurar una organización de la sociedad en función de criterios principalmente utilitaristas. (Sadin, 2020)

« El problema es hasta dónde queremos entregar nuestra libertad y qué habilidades estamos dispuestos a perder por comodidad. »



Es preciso hacerse preguntas y reflexionar críticamente sobre ello. ¿Qué desafíos presenta la IA a la filosofía política? ¿Cómo manejar los sesgos de la programación de la IA a la hora de contratar personas, evaluarlas en su trabajo o en la persecución del crimen sabiendo que alucina, se equivoca y también discrimina? ¿Cuáles serán los efectos políticos de la robótica en términos de justicia e igualdad? ¿Cuáles son los impactos que tiene sobre la democracia a la hora de manipular votantes? ¿Cómo está transformando el periodismo y la generación de noticias? ¿Cómo transforma las relaciones humanas, el aprendizaje y la salud mental? ¿Cuál debería ser el



grado de participación ciudadana en la regulación de la IA? ¿Cómo impacta en los animales y en la producción agrícola? ¿Qué efectos puede tener sobre el clima y el medioambiente? ¿Cómo serían los derechos digitales para la protección de datos y asegurar el respeto por la dignidad de las personas?

Por otra parte, confundimos predicciones con futuro, como si una nueva forma de superstición nos diera la certeza de un futuro controlable o cognoscible.

La inteligencia artificial aprende del pasado. El futuro que calcula no es un futuro en el sentido propio de la palabra. Aquella es ciega

para los acontecimientos. Pero el pensamiento tiene un carácter de acontecimiento. Pone algo distinto por completo en el mundo... La inteligencia artificial solo elige entre opciones dadas de antemano, últimamente entre el uno y el cero. No sale de lo antes dado hacia lo intransitado. (Han, 2021)

Sedentarismo cognitivo

¿Qué pasaría si le pidiéramos a alguien que hiciera ejercicios físicos por nosotros y nos librara del esfuerzo que implican tales actividades? Lo obvio: perderíamos la

posibilidad de mejorar nuestra condición física, nuestra salud, y nos iríamos atrofiando físicamente, con todas las consecuencias que no es necesario explicar. Incluso en la retórica del gimnasio a nadie le parece un exceso que se hable de *sacrificio, esfuerzo, dedicación, dejarlo todo, hasta que duela*, y que a cuanto más tiempo y más esfuerzo, lograremos mejores resultados: *No pain, no gain*. El cultivarse como persona en todas las dimensiones posibles es un imperativo presente en todo tiempo y cultura. Normalmente todos quieren ser mejores de lo que son y desarrollarse en distintos aspectos de su vida. Nada de esto nos resulta extraño. Sin embargo, convivimos con una paradoja en cuanto al cuidado y desarrollo de nuestras capacidades, porque no sucede lo mismo con el cultivo intelectual. ¿Qué pasaría si los criterios que usamos para el cultivo del cuerpo, del entrenamiento físico, se trasladaran a la vida intelectual? ¿Imaginan a un profesor hablando hoy de sacrificio, esfuerzo, dedicación...? Desde los padres hasta los colegas lo mirarían con extrañeza como a un dinosaurio. ¿Por qué?

Parece que asistimos a una atrofia del pensamiento, a una promoción de la cultura del atajo y del mínimo esfuerzo intelectual. Si alguien nos puede ahorrar el tiempo para pensar, leer, escribir, comparar, calcular, sintetizar, analizar, se lo agradecemos como si nos hiciera un gran favor. Y ahora que gracias a la inteligencia artificial generativa (IAG) podemos ahorrarnos de hacer trabajos académicos que desarrollan habilidades intelectuales, asistimos a cerebros que se irán atrofiando en capacidades fundamentales para pensar lúcidamente. Y no es que no sea útil usar colaborativamente la IAG para el estudio y el trabajo, sino que el problema es hasta dónde queremos entregar nuestra libertad y qué habilidades esta-

mos dispuestos a perder por comodidad. El riesgo sustancial es que dejemos de enseñar el valor del esfuerzo y de la concentración, del ser capaces de permanecer unas horas sentados enfocados en algo difícil con el propósito de resolverlo. ¿Cómo desarrollar tenacidad y resiliencia si al instante abandonamos para que alguien o algo lo resuelva y nos ahorre el esfuerzo?

Perder la capacidad de calcular, de mantener la atención o de realizar durante un tiempo un esfuerzo deliberado para resolver un problema difícil forma parte de un fenómeno que llamamos sedentarismo cognitivo. (Sigman y Bilinkis, 2023)

La mejor forma de hacer frente al sedentarismo cognitivo es transmitir la propia pasión por el saber y los beneficios de desarrollar habilidades intelectuales que nos permitan pensar por nosotros mismos con mayor profundidad y sin renunciar a nuestra libertad para elegir quienes queremos ser y a dónde queremos llegar.

¿Qué estamos dispuestos a perder?

En los sistemas automatizados actuales el ordenador asume con frecuencia el trabajo intelectual —observando y percibiendo, analizando y valorando, tomando incluso decisiones— que hasta hace poco era considerado terreno acotado para los humanos. La persona que maneja el ordenador ocupa el rol de un empleado de la tecnología que introduce datos, monitoriza las respuestas y busca fallos. En lugar de abrir nuevas fronteras de pensamiento y acción a sus colaboradores humanos, el software estrecha nuestra perspectiva. Cambiamos talentos sutiles y especializados por otros más rutinarios y menos distintivos. (Carr, 2014)

Día tras día podemos convertirnos en seres incapaces de escribir un correo, hacer una lista de compras, ubicarnos en nuestra propia ciudad, pensar una estrategia de empresa, escribir un mensaje, un discurso o un ensayo. Con gran entusiasmo y comodidad nos abandonamos a las siempre solícitas invitaciones: «¿qué puedo hacer por usted?», sintiéndonos al mismo tiempo mimados y servidos por la tecnología, mientras la sacralizamos como una instancia superior que lo hará casi todo por nosotros, y sabrá cómo hacerlo mejor.

¿Podemos imaginar los efectos en las psichés individuales y colectivas de estar en la posición de esperararlo todo —como si estuviéramos tirados en nuestro sofá— de sistemas con aspecto de mayordomos infinitamente superiores a nosotros? Este entorno organiza la atrofia tanto de nuestro impulso hacia la exterioridad como de nuestras facultades intelectuales... (Sadin, 2024)

Según Sadin, nos encontramos en una era donde todo buscará satisfacer en tiempo real objetivos bien definidos y donde no hay lugar para la espontaneidad, ni para actividades que sean consideradas inútiles o ineficientes.

En el trabajo, muchas veces importa más el desafío que el resultado final, el proceso y el significado de lo que hacemos, que nos llena de sentido. Por ello, en el mundo laboral, la clave para la autovaloración reside en el sentido de la propia obra y saber que producimos un impacto significativo.

Nos gustan aquellas cosas que hemos hecho, nuestras propias creaciones, por el hecho de ser nuestras y de saber el esfuerzo que requirieron. Quizá en algunos años solo una pequeña minoría acceda a

«**Todavía existe cierta ingenuidad en pensar que todo se soluciona con mayor cantidad de datos, como si la respuesta a los dramas humanos dependiera exclusivamente de manejo de información y no de una profunda reflexión sobre lo que somos y qué queremos realmente hacer con el futuro de los que vendrán.**»

esos desafíos que dan sentido a la vida. De ser así, esto podría ser uno de los mayores impactos de la ia en el mundo laboral. (Sigman y Bilinkis, 2023)

La verdad reducida a datos

La IA desarrolla un tipo de funciones de manejo de datos que nos supera ampliamente en capacidad y velocidad, pero no sustituye otro tipo de capacidades humanas que tienen que ver con cómo nos relacionamos, con el sentido de la vida, que no se resuelven con datos, estadísticas y patrones. La reducción del conocimiento a información lleva a un ingenuo optimismo sobre diversas posibilidades de la IA respecto de la vida humana.

Sea cual sea la dirección que tome el desarrollo de la IA, no podemos delegarle la responsabilidad ni la sabiduría. Todavía existe cierta ingenuidad en pensar que todo se soluciona con mayor cantidad de datos, como si la respuesta a los dramas humanos dependiera exclusivamente de

manejo de información y no de una profunda reflexión sobre lo que somos y qué queremos realmente hacer con el futuro de los que vendrán. Es claro que no podemos esquivar el progreso tecnocientífico y es deseable que pensemos responsablemente en cómo acompañamos estos procesos. Sería irresponsable caer en un determinismo que suponga que solo hay que subirse a la ola sin pensar, como si nada dependiera de nosotros más que aceptar un futuro ya programado por fuerzas incontrolables.

El futuro lo escribimos con nuestras decisiones del presente y es de celebrar que los actores políticos estén pensando en anticiparse a lo que vendrá, de modo responsable y escuchando a los que saben desde diversas disciplinas. La gobernanza de la tecnología será cada vez más un problema ineludible en la agenda política. El secuestro de la verdad por su reduccionismo a los datos convierte a la IA en un poder sagrado, en fuente segura para juzgar la realidad.

«Lo digital se erige como un órgano habilitado para peritar lo real de modo más fiable que nosotros mismos, así como para revelarnos dimensiones ocultas a nuestra conciencia» (Sadin, 2020). Se antropomorfiza a las máquinas como si ellas hicieran el mejor discernimiento, y a las que solo resta obedecer y aliviar el peso de tener que pensar. Ahorramos tiempo y esfuerzo mental, al mismo tiempo que entregamos nuestra libertad sin resistencia y aceptamos la nueva verdad sin reparos.

Si bien podemos trabajar colaborativamente y aprovechar las posibilidades de la tecnología, el mayor desafío es pensar responsablemente a qué estamos dispuestos a renunciar de nuestra condición humana por comodidad y cuáles son nuestros mínimos no negociables.

El pensamiento humano es más que cálculo y resolución de problemas. Despeja e ilumina el mundo. Hace surgir un mundo completamente diferente. La inteligencia de las máquinas entraña ante todo el peligro de que el pensamiento humano se asemeje a ella y se torne él mismo maquina. (Han, 2021)

Ciberleviatán y el poder tecnocrático

En su obra *Ciberleviatán* (2019), José María Lasalle plantea la encrucijada que tenemos delante como humanidad en ir perdiendo libertad por mayor seguridad o que, con responsabilidad política, exista un auténtico pacto que asegure la libertad de los ciudadanos, que proteja los datos y establezca nuevos derechos digitales.

Nos encontramos sumergidos en un enjambre de humanos «sin capacidad crítica y entregado al consumo de aplicaciones tecnológicas dentro de un flujo asfixiante de información que crece exponencialmente» (Lasalle, 2019).

Según este filósofo español, el liberalismo humanista tiene por vocación primera limitar el poder y ahora se enfrenta a la seducción del poder tecnológico que quiere ser omnipresente y omnisciente, sin resistencias. Asistimos a una nueva reconfiguración del poder:

Hoy, los datos que genera internet y los algoritmos matemáticos que los discriminan y organizan para nuestro consumo son un binomio de control y dominio que la técnica impone a la humanidad. Hasta el punto de que los hombres van adquiriendo la fisonomía de seres asistidos digitalmente debido, entre otras cosas, a su incapacidad para decidir por sí mismos. (Lasalle, 2019)

La fascinación con un poder ilimitado de la tecnología, entendido como inevitable e ineludible, que asegura mayor control y niveles de certeza en decisiones, va desconfiando paulatinamente de la fragilidad y espontaneidad del factor humano. Así, la libertad tan valorada y defendida, comienza a ser vista como un problema para el progreso, por lo cual los humanos deberían aceptar que su libertad sea asistida por una inteligencia superior, cuasi divina: la inteligencia artificial. Algunos autores comienzan a ver en este cambio sociocultural tecnocrático la promesa de proteger a los humanos de su peligrosa espontaneidad y, bajo una preferencia determinista, sería mejor programarnos para lo que se considere mejor según criterios utilitaristas.

Vamos perdiendo libertades con la ilusión de que *podemos acceder* a nuevas posibilidades de desarrollo, como si para ser ilusoriamente más libres tuviéramos que ir renunciando a libertades fundamentales. Y lo hacemos pasivamente y con cierta naturalidad y fascinación.

Nos encontramos así con un alineamiento entre lo técnico, lo económico y lo político donde el poder se centraliza de forma desmesurada sobre un progresivo número de actividades, incluyendo la salud, la educación y el trabajo. Según Lasalle, «el despotismo algorítmico está haciendo volver a los hombres a una nueva minoría de edad que desanda la tradición liberal del conocimiento que propició la Ilustración». ♦

Bibliografía para profundizar

- CARR, N. (2014). *Atrapados. Cómo las máquinas se apoderan de nuestras vidas*. Barcelona: Taurus.
- COECKELBERGH, M. (2021). *Ética de la inteligencia artificial*. Madrid: Cátedra.

- COECKELBERGH, M. (2023). *La filosofía política de la inteligencia artificial*. Madrid: Cátedra.
- CRAWFORD, K. (2022). *Atlas de la inteligencia artificial. Poder, política y costos planetarios*. Buenos Aires: FCE.
- HAN, B.-CH. (2021). *No-cosas: quiebres del mundo de hoy*. Buenos Aires: Penguin Random House.
- LASALLE, J. M. (2019). *Cyberleviatán: El colapso de la democracia liberal frente a la revolución digital*. Madrid: Arpa.
- SADIN, E. (2020). *La inteligencia artificial o el desafío del siglo. Anatomía de un anti-humanismo radical*. Buenos Aires: Caja Negra.
- SADIN, E. (2022). *La era del individuo tirano. El fin del mundo común*. Buenos Aires: Caja Negra.
- Sadin, E. (2024). *La vida espectral. Pensar la era del metaverso y las inteligencias artificiales generativas*. Buenos Aires: Caja Negra.
- SAVULESCU, J., Y LARA, F. (2021). *Más que humanos. Biotecnología, inteligencia artificial y ética de la mejora*. Madrid: Tecnos.
- SIGMAN, M., Y BILINKIS, S. (2023). *Artificial. La nueva inteligencia y el contorno de lo humano*. Barcelona: Penguin Random House.
- VARELA, L. (2022). *Espejos: filosofía y nuevas tecnologías*. Barcelona: Herder.



Miguel Pastorino

Doctor en Filosofía. Máster en Bioética. Magister en Dirección de Comunicación. Profesor del Departamento de Humanidades Comunicación de la Universidad Católica del Uruguay y director del Centro Core de esta.

Core de esta.

X: @MiguelPastorino

IG: @pastorinomiguel

LI: miguelpastorino



Personalizar las *fakes*: ¿hacia el apocalipsis de la desinformación?

CHRISTOPH NEHRING

Este artículo explora los efectos y las implicaciones de la desinformación generada por la IA. Examina sus formas, en particular los *deepfakes*, y su impacto en las elecciones recientes y futuras. También ofrece ideas prácticas para identificar y combatir la desinformación generada por la IA, con especial atención al papel de los *influencers*, los periodistas y otros profesionales de los medios, y a los desafíos únicos que enfrentan.

La inteligencia artificial está transformando el panorama informativo global, creando oportunidades y riesgos sin precedentes. La generación, difusión y amplificación de la desinformación y la misinformación (información errónea) son ejemplos destacados. A pesar del miedo y la confusión generalizados, el conocimiento empírico sobre la desinformación generada por IA, sus formas, impacto y efectos sigue siendo limitado, lo que alimenta la incertidumbre, el temor, la desconfianza y la demanda de información equilibrada y de calidad.

Desinformación, *deepfakes* y manipulación

Ya en 2023 los expertos en desinformación señalaban el potencial de la IA generativa como un arma de engaño masivo, que potencia y amplifica la desinformación y la misinformación. Aunque estos escenarios catastrofistas aún no se han materializado, la IA tiene varias cualidades que afectan significativamente la producción y distribución de desinformación y misinformación. La IA puede hacer que la desinformación sea:

- *más rápida*, tanto en la creación de contenidos como en su distribución automática;

- *más barata*, p. ej., automatizando la producción y la distribución, reduciendo los recursos humanos y financieros;
- *más persuasiva*, p. ej., utilizando *deepfakes* superrealistas;
- *más personalizada*, p. ej., utilizando IA para analizar datos y ajustar los mensajes para alcanzar a públicos específicos;
- *de mayor alcance*, p. ej., usando bots de IA y automatizando la distribución de desinformación, o simplemente porque las herramientas de IA están disponibles para todos los usuarios de redes sociales.

Experimentos de hackers y periodistas muestran en un ejemplo que el costo de usar ChatGPT para operar una web de noticias falsas totalmente automatizada bajó de USD 400 a USD 105 entre 2023 y 2024.

Formas de desinformación de las IA

La revolución de la GenAI impacta en el software que crea contenidos, textos, imágenes, videos y audio. Las formas conocidas de desinformación generada por IA incluyen:

- a. *Sitios web de noticias falsas.* Aunque difíciles de detectar, se han identificado miles de sitios cuyos contenidos (texto, imágenes, videos) son generados por ChatGPT u otros chatbots. Algunos, como Electionwatch o TheDCWeekly, se enfocan en la desinformación organizada sobre la política estadounidense y las elecciones presidenciales de 2024, mientras que otros son sitios comerciales que reescriben y publican noticias antiguas con fines lucrativos.
- b. *Imágenes de IA.* Las imágenes generadas por IA están inundando redes sociales, servicios de mensajería y sitios web. Algunos muestran a personas, generalmente políticos, en situaciones ficticias (como Donald Trump bailando con menores) o describen eventos que nunca ocurrieron (como un atentado terrorista al Pentágono). Si bien algunas páginas profesionales de noticias falsas, probablemente respaldadas por actores estatales, usan imágenes generadas por IA junto con artículos falsos, la mayoría de estas imágenes son creadas y compartidas por usuarios comunes en redes sociales y foros. Este tipo de imágenes se difunden especialmente durante el conflicto en Oriente Próximo, destacando los daños de la guerra y las víctimas en Gaza. En algunos casos, las imágenes de IA difundidas en redes sociales llegaron a bases de datos de stock (p. ej., Adobe Stock) donde se vendieron para uso comercial. Por otro lado, los usuarios ucranianos recurren cada vez más a imágenes generadas por IA para expresar apoyo al ejército ucraniano en la guerra contra Rusia. Esta tendencia ilustra

los efectos de la democratización de las herramientas genai y su uso indebido.

- c. *Deepfakes.* Los llamados *deepfakes* (derivados de *deep learning* y *fakes*) son contenidos de video y audio producidos o manipulados por IA. Existen varios tipos de *deepfakes*, que varían según su aplicación (p. ej., el intercambio de caras en la pornografía o en llamadas fraudulentas) o la intención detrás de ella. Los *deepfakes* creados para desinformación política han aparecido en diversos contextos, como la guerra rusa contra Ucrania y, en particular, durante las campañas electorales en todo el mundo (ver más abajo). En la mayoría de los casos se utilizan para crear pruebas falsas que desacreditan declaraciones o posturas, o para representar participación en actos ilegales o pornografía. Sus víctimas suelen ser personas públicas, como celebridades, políticos, CEO, *influencers* y periodistas. Los *deepfakes* generan un alto nivel de temor y confusión en el público debido a: a) la impresionante calidad de las falsificaciones; b) su capacidad para convencer y persuadir al público; c) la falta de software y métodos de detección fiables; y d) la inseguridad e incapacidad del público para reconocer y enfrentar estos contenidos. En las siguientes secciones de este ensayo se abordarán los *deepfakes* como uno de los ejemplos más urgentes y peligrosos de desinformación generada por IA.

Expertos e investigadores estatales han encontrado pruebas empíricas de la existencia de todas estas formas de desinformación generada por IA. Sin embargo, el llamado *detection challenge* ‘reto en la detección’ del

contenido de IA dificulta evaluar y determinar con precisión la cantidad y el alcance real de la desinformación generada por IA. Hasta la fecha no existe un método de detección cien por ciento preciso para contenidos generados por IA, ni filtros automáticos de carga ni servicios de eliminación, etc. Esto significa que, aunque la calidad y cantidad de la desinformación generada por IA están aumentando rápidamente, su alcance real sigue siendo difícil de evaluar.

Deepfakes y elecciones en 2023 y 2024

En los últimos dos años, la desinformación de IA, especialmente los *deepfakes*, se ha convertido en un arma para influir en campañas políticas y elecciones. En la mayoría de los casos, la tecnología se ha usado para crear videos o audios de políticos, candidatos, periodistas y otras figuras públicas en situaciones negativas y desacreditadoras. Algunos atacan la reputación de individuos para socavar su credibilidad, imagen y reputación pública, mientras que otros son parte de campañas políticas negativas que buscan desacreditar opiniones, decisiones o eventos políticos. Todos ellos intentan influir en el comportamiento de los votantes, difundiendo deliberadamente información falsa, engañosa o descontextualizada, creada artificialmente.

En otros casos, los *deepfakes* se usan en campañas políticas oficiales. Estos *deepfakes* se distinguen porque: a) están vinculados a una fuente *oficial* como un candidato, partido, institución u organización; b) están (frecuentemente) etiquetados como contenido generado por IA; y c) no contienen necesariamente información falsa. Durante las elecciones al Parlamento Europeo de junio de 2024, varios partidos de extrema derecha y derecha (p. ej., Francia e Italia) usa-

» La desinformación de IA, especialmente los *deepfakes*, se ha convertido en un arma para influir en campañas políticas y elecciones. «

ron tecnología para promover sus mensajes y narrativas mediante memes, imágenes y canciones generadas por IA. En *Pakistán*, el ex primer ministro Imran Khan y su equipo usaron *deepfakes* para incluirlo en videos de campaña mientras estaba encarcelado. En *India*, *Indonesia* y *Filipinas*, partidos y equipos de campañas crearon *deepfakes* de políticos fallecidos o figuras públicas populares para sus campañas electorales. Durante las elecciones presidenciales en *Argentina*, ambos candidatos y sus equipos utilizaron extensamente todas las formas de IA generativa (imágenes, videos, texto) en sus campañas. Esto incluyó también videos malintencionados de ambos candidatos, que cruzaron la línea entre la campaña y la desinformación al difundir deliberadamente mentiras provocadoras. En *México*, la entonces candidata presidencial y ex jefa de Gobierno de Ciudad de México, Claudia Sheinbaum, apareció en un video que supuestamente promovía un esquema financiero fraudulento, lo que dañó su credibilidad política. En todos los países y todas las elecciones de 2024, se vieron *deepfakes* políticos diseñados para desacreditar a candidatos y promover narrativas, mayormente agresivas. Durante las elecciones presidenciales de Estados Unidos, el uso de IA generativa fue especialmente prominente. Ambas partes —canales oficiales y simpatizantes— publicaron imágenes generadas por IA para transmitir sus

mensajes. Sin embargo, en redes sociales circularon formas más peligrosas de *deepfakes*, como *robollamadas* con la voz del presidente Joe Biden instando a no votar, *AI-fakes* mostrando a Taylor Swift apoyando a Donald Trump, y contenidos falsos creados por IA que supuestamente aparecían en el libro de J. D. Vance.

A pesar de que los *deepfakes* estuvieron presentes en todas las elecciones de 2024, las pruebas empíricas sugieren que, contrariamente a las expectativas apocalípticas, no tuvieron un impacto significativo en los resultados. Hasta ahora, solo en dos casos los *deepfakes* ocurridos en las 48 horas previas a las elecciones han tenido una influencia decisiva. En Eslovaquia, un de audio en el que un candidato presuntamente discutía la compra de votos de minorías pareció afectar directamente los resultados, aunque no inclinó los resultados finales a favor de otro candidato. Por otra parte, durante las elecciones presidenciales en Turquía, un video *deepfake* de contenido pornográfico que involucraba a uno de los candidatos resultó en su retirada de la contienda electoral. Obviamente, esto influyó en el resultado de las elecciones, pero dado que todas las encuestas ya apuntaban al presidente en funciones como claro vencedor, el *deepfake* puede haber tenido algún impacto, aunque no llegó a inclinar el resultado final.

Periodismo e influencers: el espacio informativo mundial

La IA generativa tiene el potencial de transformar completamente el panorama informativo global y afectar todas las formas de comunicación política, creación y presentación de contenidos, incluyendo el periodismo y el trabajo de los *influencers*.

IA y periodismo

La GenAI tiene un impacto significativo en la creación y presentación de contenidos en el ámbito del periodismo. Sin embargo, existe una aparente brecha de la IA: mientras los medios de comunicación tradicionales y de calidad se esfuerzan por encontrar respuestas, establecer límites y regular el uso ético de la IA en el periodismo, los medios de baja calidad, los tabloides, los medios de propaganda estatal y los estafadores ya están aprovechando la IA para sus propios fines. El medio ruso de propaganda extranjera RT, por ejemplo, ya utiliza personajes *deepfake*, es decir, avatares automatizados generados completamente por IA, a los que llaman presentadores digitales, para su programa en español. También se sabe desde hace tiempo que varios canales de noticias de China y otros países han estado utilizando la IA con estos fines.

Mientras que los medios tradicionales de calidad de todo el mundo suelen abstenerse de utilizar GenAI para crear noticias centrales y continúan haciéndolo por sí mismos, otros actores son más propensos a emplearla. Las agencias de noticias han identificado miles de sitios web que utilizan IA (principalmente ChatGPT) para gestionar sitios de noticias totalmente automatizados. Estos sitios web se dedican tanto a republicar y reescribir contenidos antiguos para generar ingresos publicitarios, como a difundir directamente desinformación política. Por otro lado, Channel 1, una nueva emisora de noticias establecida en Los Ángeles en 2024, es el primer medio que afirma ser un actor serio en el periodismo pero que gestiona íntegramente sus programas con GenAI, tanto en la creación de contenidos como en su presentación.

Otro aspecto importante de GenAI en el periodismo es cómo las platafor-



mas de redes sociales regulan, marcan y publican contenidos generados por IA. Aunque pronto será obligatorio que las plataformas marquen y especifiquen los contenidos generados por IA en América del Norte y la UE, no existen normas unificadas de este tipo en otras partes del mundo. La mayoría de las plataformas sociales afirman en sus normas comunitarias y condiciones de uso que los contenidos y perfiles generados por IA deben estar claramente marcados y registrados. Sin embargo, como en el pasado, el grado de cumplimiento de estas normas varía considerablemente.

IA e influencers

En el mundo de los *influencers* parece estar ocurriendo una evolución similar con la generación y presentación de contenidos, que son profundamente impactadas por GenAI. Los *influencers* virtuales, es decir, avatares creados y gestionados completamente por IA que se hacen pasar por *influencers* en plataformas de redes sociales, ya han atraído a millones de seguidores en países como China, Brasil, Estados Unidos e India. Esto también impacta la difusión de desinformación, desinformación, teorías conspirativas, etc. En todo el mundo,

los *influencers* están adquiriendo cada vez más relevancia, en tanto grupo objetivo y herramienta para los actores profesionales de la desinformación, así como creadores y difusores de desinformación.

Algunos *influencers* de TikTok, por ejemplo, han convertido en su modelo de negocio la creación de videos generados por IA sobre nuevas teorías conspirativas, y debaten en chats cerrados sobre cómo utilizar GenAI para incrementar sus ingresos. En otros casos, se ha demostrado que embajadas rusas en África pagan a *influencers* locales para difundir desinformación. Hasta ahora, no se ha comprobado que los *influencers* generados por IA estén involucrados en campañas políticas o en la difusión de desinformación, aunque el riesgo sigue siendo alto. Las manipulaciones de IA y los *deepfakes* también afectan a los profesionales de los medios y a los *influencers*: ambos grupos son frecuentemente víctimas de ataques de desprestigio mediante estas tecnologías. Los videos *deepfake* en los que periodistas promocionan, sin su consentimiento, estafas financieras o productos dudosos se han vuelto comunes en Estados Unidos y Europa. Los *influencers*, por su parte, a menudo enfrentan el riesgo de ser víctimas de *deepfakes* que dañan su reputación (y, en consecuencia, su modelo de negocio). El escenario más común en este caso es el uso de imágenes de *influencers* mujeres en *deepfakes* de contenido por-

» Rusia es considerada uno de los actores más activos en la manipulación e injerencia de información extranjera. «

nográfico. Estos ataques también pueden ser parte de campañas de desprestigio con fines políticos, como se vio con los *deepfakes* de Taylor Swift tras su involucramiento en elecciones estadounidenses de 2024.





Sobrecontaminación: ahogándonos en un mar de contenidos generados por IA

Otra dimensión de GenAI en el ámbito global de la información es la posibilidad real de su sobrecontaminación con contenidos

» La inteligencia artificial está transformando de manera radical el panorama de la desinformación, la interferencia electoral y la manipulación de la información. «

generados por IA, robots automatizados y similares. Los escenarios pesimistas sugieren que, para 2026, el noventa por ciento de todos los contenidos en línea podrían ser generados por IA. Además, algunos estudios indican que el comportamiento automatizado en línea, como bots y programas, ya constituye la mayoría de las actividades en la red. Si la GenAI llega a dominar la mayor parte del contenido, la presentación y las actividades en línea, esto impactará gravemente en las noticias políticas y en la información en general y afectará a las sociedades. Por lo tanto, la sobrecontaminación podría convertirse en uno de los riesgos más serios de la GenAI a largo plazo.

Manipulación e injerencia de información extranjera rusa

Rusia es considerada uno de los actores más activos en la manipulación e injerencia de información extranjera (FIMI). La difusión coordinada y encubierta de información falsa, engañosa y manipulada es una herramienta clave para influir en sociedades, eventos y elecciones en estas actividades. La desinformación rusa está presente en casi todas las elecciones del mundo y utiliza una amplia gama de herramientas e instrumentos complejos. Las

embajadas y consulados rusos, medios de comunicación, empresas de relaciones públicas, periodistas *freelance*, *influencers*, portales web anónimos y proxis locales son los principales actores en la desinformación rusa. Sus tácticas abarcan desde la propaganda básica y el pago a *influencers* y periodistas hasta operaciones de desinformación más complejas que incluyen la falsificación de medios de comunicación respetados y la difusión encubierta de noticias falsas. Las narrativas y mensajes de la desinformación rusa suelen centrarse en temas clave (p. ej., antioccidentalismo, anti-Ucrania y anti-LGTBQ) que se adaptan en mensajes personalizados para audiencias locales en diferentes partes del mundo. En el Sur Global, estas narrativas suelen enfocarse en desacreditar a Occidente (p. ej., colonialismo, tensiones sociales, injusticia económica y social).

Usos de la IA

En sus operaciones de desinformación, los actores rusos de la FIMI han demostrado su disposición a explotar el potencial ilimitado de la GenAI para sus objetivos. El programa en español de la emisora rusa de propaganda extranjera RT ahora incluye dos presentadores digitales, es decir, avatares generados por IA. En Estados Unidos, varios sitios web de noticias falsas que publicaron artículos negativos completamente automatizados sobre las elecciones presidenciales, escritos con GenAI, tienen vínculos con Rusia. Durante una sofisticada campaña mundial de desinformación llamada *Doppelganger*, que se enfoca en sitios web falsificados de los medios tradicionales más conocidos, se descubrió que agentes rusos usaban ChatGPT para generar y traducir publicaciones y comentarios en redes sociales.

❖ Las estrategias basadas en IA pueden utilizarse para dirigir mensajes a votantes específicos, personalizar el contenido y aumentar la eficacia de las campañas políticas. ❖

En la guerra en curso contra Ucrania, los actores rusos han utilizado repetidamente videos *deepfake* (p. ej., uno falso del presidente Zelensky pidiendo la rendición y otro fabricado de los jefes de inteligencia ucranianos supuestamente admitiendo su participación en un atentado terrorista en Moscú) para desinformar tanto a la audiencia interna como externa.

Conclusiones

La inteligencia artificial está transformando de manera radical el panorama de la desinformación, la interferencia electoral y la manipulación de la información. Hoy en día, todas estas actividades malintencionadas han integrado la IA, y ya no hay elecciones sin algún nivel de desinformación generada por esta tecnología. La IA facilita la producción y difusión de este tipo de contenido a una velocidad sin precedentes, reduciendo costos y simplificando su creación. Esto ha permitido que las campañas de desinformación sean más accesibles, automatizadas, personalizables, persuasivas y de mayor alcance.

A pesar de estos avances, el temido *apocalipsis de la información* aún no se ha materializado. Ninguna elección ha sido

influida de forma decisiva por la desinformación generada por la IA, aunque ha habido casos notables en los que los *deepfakes* han jugado un papel significativo. Por ejemplo, en las recientes elecciones en Turquía y Eslovaquia, aunque los *deepfakes* llamaron la atención y suscitaron preocupación, al final no alteraron los resultados a favor de ningún candidato o partido.

Mientras tanto, la IA no es solo una herramienta para la desinformación, sino también una fuerza creciente en las campañas políticas. Las estrategias basadas en IA pueden utilizarse para dirigir mensajes a votantes específicos, personalizar el contenido y aumentar la eficacia de las campañas políticas. A medida que esta tendencia crece, también aumentan los riesgos asociados a las manipulaciones con IA, especialmente los *deepfakes*. Más allá de las elecciones, los *deepfakes* alimentan el ciberacoso, el fraude, las estafas y las violaciones de ciberseguridad, con los *influencers* siendo particularmente vulnerables a estos usos maliciosos de la tecnología de IA, como el ciberacoso mediante pornografía *deepfake*. ♦

Bibliografía

- BONTCHEVA, K. (ed.) (2024). Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities. <https://www.veraai.eu/posts/white-paper-generative-ai-and-disinformation>
- FERRARA, E. (2024). GENAI Against Humanity: Nefarious Applications of Generative Artificial Intelligence and Large Language Models. <https://doi.org/10.48550/arXiv.2310.00737>
- GEHRINGER, F. A., NEHRING, CH., Y LABUZ, M. (2024, 10 de mayo). The influence of Deep Fakes on Elections: Legitimate Concern or Mere Alarmism? *KAS Monitor* 2024. <https://www.kas.de/en/monitor/detail/-/content/the-influence-of-deep-fakes-on-elections>
- HABGOOD-COOTE, J. (2023). Deepfakes and the epistemic apocalypse. *Synthese*, v. 201 103/2023. <https://link.springer.com/article/10.1007/s11229-023-04097-3>
- LABUZ, M., NEHRING, CH. (2024, 26 de abril). On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *Eur Polit Sci*. <https://doi.org/10.1057/s41304-024-00482-9>
- MARCHAL, N., Y XU, R. (2024, 2 de agosto). Mapping the misuse of generative AI. *GoogleDeepmind*. <https://deepmind.google/discover/blog/mapping-the-misuse-of-generative-ai/>
- MUÑOZ, M. (2024). *The AI Election Year: How to Counter the Impact of Artificial Intelligence*. DGAP Memo, v. 1. <https://dgap.org/en/research/publications/ai-election-year>
- SCHICK, N. (2020). *DEEP FAKES AND THE Infocalypse*. Ottawa.

Traducción inglés-español: Doris Filipovic.



Christoph Nehring

Investigador, analista y periodista. Profesor invitado y analista en el programa de medios de comunicación de la Fundación Konrad Adenauer, autor para *Tagespiegel*, *Deutsche Welle*, *nzz*, *Spiegel* y muchos otros. Apasionado de la IA y la desinformación. Lleva más de diez años investigando la desinformación, la manipulación y los servicios secretos.

LI: christopher-nehring-423b06257

DIÁLOGO POLÍTICO es una plataforma para el diálogo democrático entre los influenciadores políticos sobre temas de relevancia en América Latina con base en los valores de libertad, solidaridad y justicia. Conecta a la región con los grandes debates geoestratégicos en el mundo. Construye una ventana de difusión de proyectos de la Fundación Konrad Adenauer en América Latina.

DIÁLOGO POLÍTICO es parte del Programa Regional Partidos Políticos y Democracia en América Latina (KAS Partidos). Tiene el objetivo de reducir la polarización política a través de un debate pluralista, constructivo e informado, orientado al bien común, para fortalecer el centro político desde sus raíces socialcristianas, liberales y conservadoras.

-
- | | | | |
|-----------|--|------------|---|
| 6 | Prólogo
Democracia artificial | 66 | La lucha anticorrupción desde el nuevo paradigma tecnológico
<i>Denisse Rodríguez-Olivari</i> |
| 8 | Inteligencia artificial y elecciones: ¿amenazas prematuras?
<i>Sarah Kreps</i> | 70 | IA y seguridad pública en las democracias latinoamericanas
<i>Gustavo Flores Macías y Bárbara Hernández</i> |
| 22 | El gobierno de los algoritmos
<i>Daniel Innerarity</i> | 82 | Lars Zimmermann: «La tecnología no nos eximirá de las tareas de la democracia»
<i>Entrevista</i> |
| 30 | Gobernanza anticipatoria con ayuda de la tecnología
<i>Lydia Garrido Luzardo</i> | 92 | La libertad y la verdad secuestradas por la inteligencia artificial
<i>Miguel Pastorino</i> |
| 40 | ¿Es posible regular la IA? Experiencias a nivel global
<i>Ximena Docarmo</i> | 102 | Personalizar las fakes: ¿hacia el apocalipsis de la desinformación?
<i>Christoph Nehring</i> |
| 50 | De las redes masivas al voto personalizado
<i>Jesús Delgado Valery</i> | | |