

# 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 (Martinou, 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.

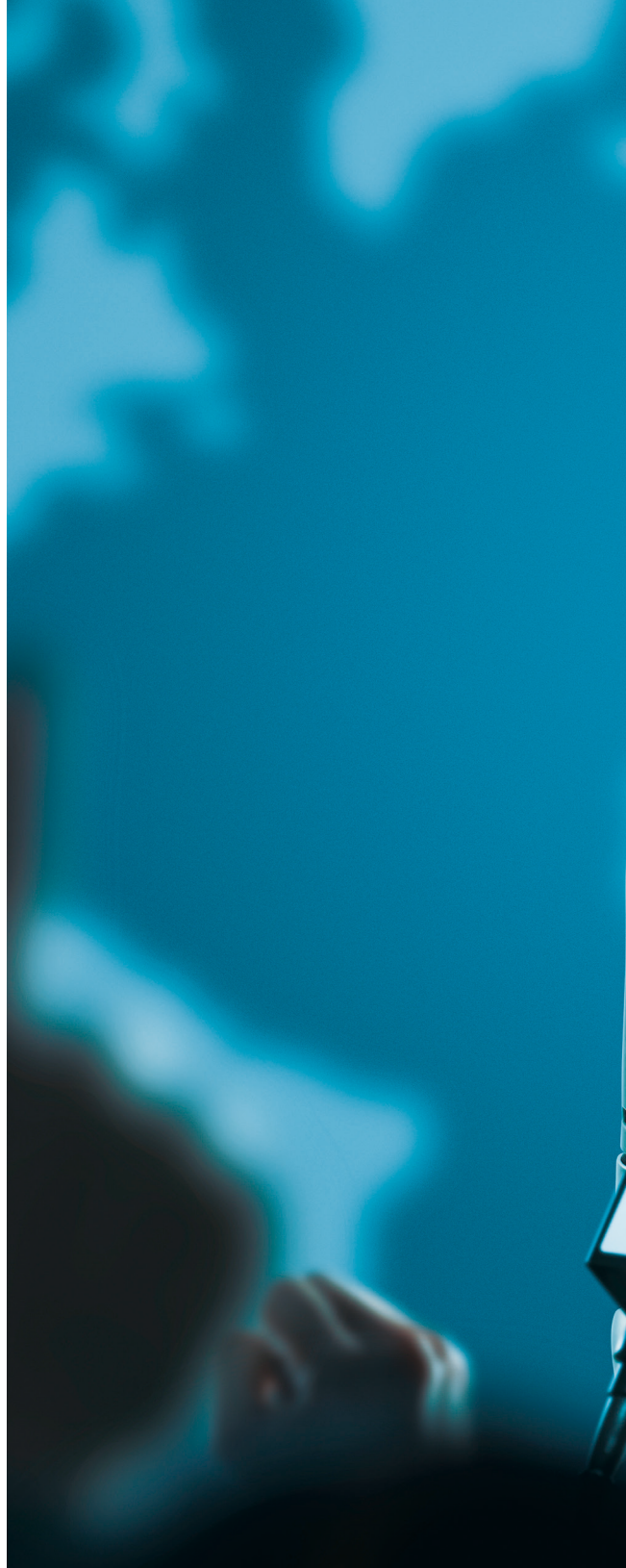
En Eslovaquia, un audio atribuyó falsamente planes para amañar unas elecciones a un líder político antes de las elecciones

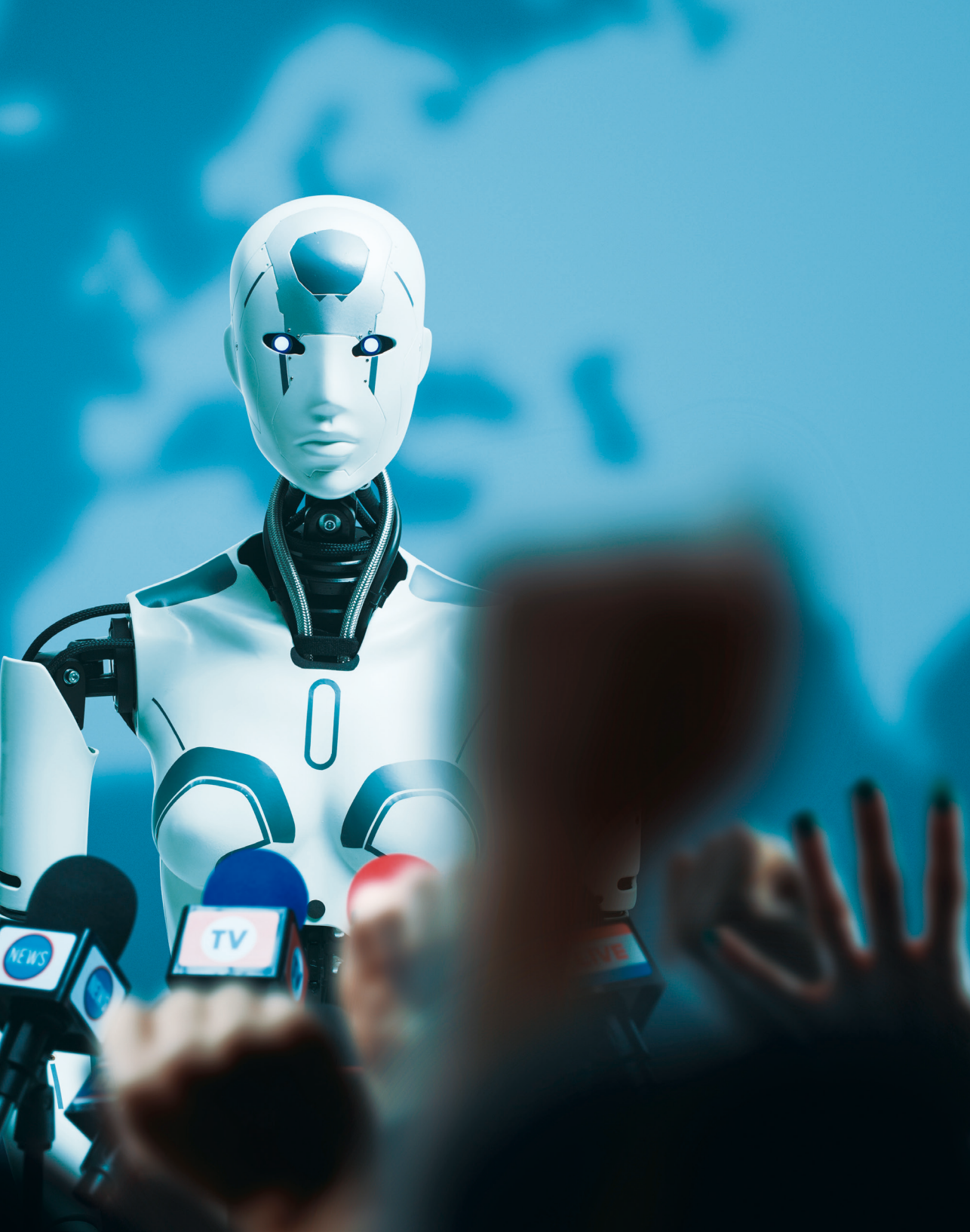
parlamentarias de 2023 (Devine, 2024). También se han utilizado imágenes y videos generados por IA para tergiversar la imagen de los políticos. Imágenes generadas por IA mostraban a Donald Trump con votantes negros con la intención de ganarse su simpatía (Sainato, 2024). Durante las primarias de 2023, 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](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/](http://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