International IDEA

FGV COMUNICAÇÃO

INTELIGENCIA ARTIFICIAL E INTEGRIDAD DE LA INFORMACIÓN: EXPERIENCIAS LATINOAMERICANAS

Documento de Política Nº 34, 2025

INTELIGENCIA ARTIFICIAL E INTEGRIDAD DE LA INFORMACIÓN: EXPERIENCIAS LATINOAMERICANAS

Documento de Política Nº 34, 2025

Sebastián Becker Castellaro, Mariana Carvalho, Alberto Fernández Gibaja, Amaro Grassi, Cecilia Hammar, Juliane Muller, Laura Pereira, Victor Piaia, Marco Aurélio Ruediger

IDEA Internacional
Strömsborg
SE-103 34 Estocolmo
SUECIA
+46 8 698 37 00
info@idea.int
www.idea.int

© 2025 Fundação Getulio Vargas—Comunicação y Instituto Internacional para la Democracia y la Asistencia Electoral

Las publicaciones del Instituto Internacional para la Democracia y la Asistencia Electoral (IDEA Internacional) son independientes de intereses específicos nacionales o políticos. Las opiniones expresadas en esta publicación no representan necesariamente las opiniones de IDEA Internacional, de su Junta Directiva ni de los Miembros de su Consejo y Fundação Getulio Vargas—Comunicação.

© 2025 Edición en inglés

Con la excepción de las imágenes y fotografías de terceros, la versión electrónica de esta publicación está disponible bajo licencia de Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0). Se permite copiar, distribuir y transmitir esta publicación, así como usarla y adaptarla, siempre que sea únicamente para fines no comerciales, se reconozca adecuadamente la publicación y se distribuya bajo una licencia idéntica. Para obtener más información sobre esta licencia, consulte el sitio web de Creative Commons: <http://creativecommons.org/licenses/by-nc-sa/4.0>.

IDEA Internacional
Strömsborg
SE-103 34 Estocolmo
SUECIA
Teléfono: +46 8 698 37 00
Correo electrónico: info@idea.int
Sitio web: <https://www.idea.int>

Imagen de portada: Generado con ChatGPT
Diseño: IDEA Internacional
Editora: Pablo Batalla Cueto
Traduccion: con IA

DOI: <https://doi.org/10.31752/idea.2025.44>

ISBN: 978-91-7671-976-3 (versión en pdf)

Agradecimientos

Este documento ha sido desarrollado conjuntamente por el Programa de Digitalización y Democracia de IDEA Internacional y la Fundação Getulio Vargas-Comunicação.

Agradecemos a Enzo Martino por su ayuda en la investigación; a Nicole Ingouville de Souza y Victor Rabello Piaia por coordinar la realización de este documento; y a Katarzyna Gardapkhadze, Nicolás Liendo y Marcela Ríos Tobar por su revisión y aportes en varios borradores. También estamos agradecidos a Lisa Hagman por liderar la producción del ensayo.

Abreviaturas

DaaS	Desinformación como servicio
IA	Inteligencia artificial
INE	Instituto Nacional Electoral, México
LGBTQIA+	Lesbianas, gays, bisexuales, transgénero, queer, intersexuales y asexuales
LLMs	Modelos extensos de lenguaje [Large language models]

Índice

Agradecimientos ..... **iv**

Abreviaturas ..... **v**

Resumen ejecutivo **1**

Introducción: El desafío de la inteligencia artificial en América Latina **3**

Capítulo 1
La IA y la integridad de la información en las elecciones **7**
1.1. La IA y la creación y distribución de información en América Latina 8
1.2. La IA y los grupos marginalizados 16

Capítulo 2
Impacto de la inteligencia artificial en las elecciones de México y Brasil **20**
2.1. Uso de la inteligencia artificial generativa en las elecciones nacionales y locales 22
2.2. La IA como amplificador de la violencia política de género 26
2.3. Herramientas para la promoción de la ciudadanía 26

Capítulo 3
Políticas y elecciones de IA en América Latina **30**
3.1. Marco analítico 30
3.2. Análisis global 32
3.3. Análisis regional: América Latina 35

Capítulo 4
Recomendaciones de política **40**

Referencias **43**

Sobre los autores **51**

Acerca de IDEA Internacional **53**

FGV Comunicação **54**

RESUMEN EJECUTIVO

La inteligencia artificial (IA) podría remodelar el panorama electoral de América Latina al exacerbar aún más los problemas existentes por la forma en que se crea, selecciona y difunde la información. Los avances en el procesamiento del lenguaje natural, arquitectura basada en transformers [*transformer-based architecture*] y el acceso generalizado a grandes conjuntos de datos han dado lugar a potentes aplicaciones de IA. Si bien la IA mantiene la promesa de un fomento de la participación cívica, el impulso del desarrollo económico y una mejora potencial de los procesos democráticos, también presenta riesgos profundos para la integridad de la información, especialmente en lo que respecta a las elecciones.

Este documento analiza el creciente papel de la IA en procesos electorales latinoamericanos, centrándose especialmente en sus implicaciones para las elecciones y los grupos marginados. Analiza los usos de la IA en cuatro niveles (creación, curación y moderación de contenidos, así como publicidad política) y documenta su presencia desigual pero en expansión en las elecciones de 2024 en Brasil y México. Los estudios de caso de las elecciones en ambos países revelan la utilización de la IA generativa tanto para engañar como para convencer, incluso a través de *deepfakes,* contenido de audio manipulado y *chatbots* impulsados por IA. Especialmente preocupante es el uso de la IA para amplificar la violencia política de género. En tanto las candidatas femeninas y LGBTQIA+ son objeto de ataques desproporcionados a través de contenidos no consentidos y degradantes.

El informe también analiza los marcos de políticas regionales y mundiales en materia de IA. El estudio concluye que, si bien muchos de estos marcos reconocen la necesidad de supervisión y transparencia humanas, la mayoría descuida áreas clave: la manipulación electoral impulsada por la IA, la amplificación algorítmica de contenido dañino, los perjuicios de género y la monetización de la desinformación. Las estrategias de IA en América Latina suelen ser difusas o silenciosas, especialmente cuando se trata de

las consecuencias electorales de la IA y su impacto en las comunidades marginadas.

Para mitigar los riesgos asociados al uso de inteligencia artificial en los procesos electorales, el documento recomienda que las plataformas aumenten su transparencia. Esto incluye divulgar los conjuntos de datos utilizados para entrenar los algoritmos y explicar por qué cierto contenido se muestra a determinados usuarios. Asimismo, se sugiere promover valores democráticos, como el pluralismo mediático y la diversidad de opiniones, al tiempo que se limitan las prácticas abusivas en el uso de datos. Por otro lado, clarificar los estándares editoriales aplicables al contenido generado por IA y permitir auditorías independientes sobre los datos de las plataformas puede contribuir a evitar una amplificación algorítmica no intencionada o perjudicial. Finalmente, renovar la confianza pública en las instituciones democráticas y fortalecer el Estado de derecho son condiciones esenciales para proteger los procesos electorales frente a los efectos negativos de la IA sobre la integridad de la información, y para garantizar que todas las voces—en especial las de comunidades marginadas—sean escuchadas y representadas.

La influencia actual de la IA sobre las elecciones en América Latina sigue siendo limitada, pero su potencial disruptivo es real y creciente. El fortalecimiento de las instituciones democráticas, la mejora de los marcos regulatorios y la equidad en la gobernanza digital son cuestiones esenciales para proteger la integridad electoral en la era de la IA.

INTRODUCCIÓN: EL DESAFÍO DE LA INTELIGENCIA ARTIFICIAL EN AMÉRICA LATINA

La integridad del entorno informativo ha sido una cuestión decisiva para las elecciones en América Latina —y a nivel mundial— en los últimos años. Los actores políticos han buscado, con mayor o menor éxito, distorsionar el ecosistema de información durante las elecciones para adquirir o retener el poder. Tales actores han convertido las divisiones sociales en un arma, explotando los nuevos canales y posibilidades creados por las plataformas digitales. Han inundado la esfera pública digital con información y relatos, utilizando a menudo técnicas no auténticas o ilícitas, para promover sus intereses, erosionando la confianza en el proceso electoral. Al hacerlo, han mermado la capacidad de las elecciones para captar con precisión las preocupaciones y opiniones de los ciudadanos y transformarlas en políticas —una queja de larga data en toda América Latina—, lo que está alimentando una creciente desconfianza y desvinculación de la democracia. De Argentina a México y de Brasil a El Salvador, la contaminación del entorno informativo representa un desafío definitorio para la democracia en todo el continente.

Los actores políticos han buscado, con mayor o menor éxito, distorsionar el ecosistema de información durante las elecciones para adquirir o retener el poder.

El problema va más allá de la difusión de información falsa o inexacta: implica la manipulación del ecosistema informativo para amplificar ciertos discursos políticos, dirigir la agenda y el debate político, y normalizar o legitimar ideas y políticas que, hasta hace poco, quedaban al margen del discurso político. Esta manipulación crea realidades a medida para ciertos grupos de ciudadanos; realidades que no se basan en hechos ni en un intercambio pacífico de ideas, sino en información y relatos manipulados que benefician a ciertos sectores políticos. En América Latina, la subversión del ecosistema de la información se ha basado en la falta de confianza en los medios de comunicación, en parte como consecuencia de la concentración de la propiedad de los medios, y en la creciente hostilidad a la que se enfrentan los periodistas (*Mont'Alverne* et al., 2021). La combinación de estos fenómenos fractura las sociedades al socavar un requisito fundamental para el funcionamiento democrático: la existencia de una visión compartida de la realidad, basada en información fiable.

Las elecciones no pueden cumplir su función sin una base común de hechos. Los candidatos y sus campañas han explotado estratégicamente la erosión de esa realidad compartida para cuestionar tanto la legitimidad de los resultados electorales como la fiabilidad de los organismos responsables de administrarlos. En todo el continente, muchos actores políticos han intentado construir una realidad alternativa para sus seguidores, en la que solo ellos se presentan como una fuerza política legítima, mientras que cualquier alternativa es etiquetada como antidemocrática. Esto, irónicamente, suele manifestarse en intentos de deslegitimar el propio proceso electoral. Al minar la integridad electoral, estos actores socavan la legitimidad del gobierno y profundizan la fragmentación social. Además, el debate político sobre cómo enfrentar los desafíos que se avecinan en la región—desde el cambio climático hasta la desigualdad o la corrupción—sigue siendo confuso y carece de una dirección clara para avanzar.

Las consecuencias se han dejado sentir en toda la región. En Brasil, años de ataques coordinados contra el sistema electoral culminaron en un fallido golpe de Estado tras las elecciones generales de 2022 (Souza, 2024). En Perú, el cuestionamiento infundado de los resultados de las elecciones generales de 2021 provocó días de protestas y ahondó la crisis de confianza en el sistema. En toda la región, los resultados electorales son frecuentemente cuestionados, siguiendo una tendencia global (IDEA Internacional, 2024). Los líderes autoritarios han ascendido o consolidado su poder en países como El Salvador, Nicaragua y Venezuela.

Aunque todavía se halle en sus inicios, la IA tiene potencial de convertirse en una herramienta poderosa para aquellos que buscan distorsionar el panorama de la información.

En este sombrío contexto, el rápido auge de la inteligencia artificial ha introducido nuevas y complejas amenazas contra la ya erosionada integridad de la información durante las elecciones. Aunque todavía se halle en sus inicios, la IA, en particular la IA generativa, tiene potencial de convertirse en una herramienta poderosa para aquellos que buscan distorsionar el panorama de la información. Como se observa en los estudios de caso de Brasil y México, a medida que la IA generativa mejora, se vuelve cada vez más capaz de producir texto, imágenes, audio y video engañosos, lo que debilita aún más la fortaleza de una visión compartida entre los ciudadanos. La IA no solo puede convencer a los ciudadanos de una realidad alternativa, sino también distribuirla directamente a las pantallas de sus teléfonos. De manera alarmante, todavía no existe un método fiable para detectar contenido generado por IA con la velocidad y precisión necesarias para mitigar sus efectos más dañinos.

En tiempos de alta tensión, como unas elecciones, el peligro de este escenario es evidente. Como hemos visto en la región, las tensiones sociales durante los días decisivos de los comicios pueden verse exacerbadas por la manipulación de la información y la IA puede aumentar aún más esta tensión. El contenido generado por IA puede alimentar la controversia en torno a los resultados y la violencia política. Un gobierno electo puede ser presentado como ilegítimo o, en el peor de los casos, todo el sistema democrático verse desestabilizado. El intento de golpe de Estado en Brasil es un duro recordatorio de la

vulnerabilidad del sistema cuando se utiliza la información como arma para socavar los resultados electorales.

Sin embargo, en su estado actual y sobre la base de las experiencias de 2024, la IA parece estar mejorando solo marginalmente la capacidad de los actores políticos para distorsionar el entorno informativo en torno a las elecciones. Su uso parece centrarse en incrementar la confusión y reforzar los discursos existentes, en un esfuerzo por erosionar la confianza en el sistema. Sin embargo, no es la fuerza motriz detrás de la creación de realidades alternativas. En cualquier caso, su influencia potencial es lo suficientemente significativa como para justificar un análisis. Finalmente, es imperativo establecer salvaguardas dentro del sistema para evitar sus consecuencias más perjudiciales.

Como ejemplifican los estudios de caso de Brasil y México, las elecciones de 2024 en la región y a nivel mundial solo han mostrado usos aislados y limitados de la tecnología, mientras que sigue habiendo un potencial sustancial para que la IA tenga un impacto más amplio en todo el ciclo electoral.

Dos razones principales explican lo limitado hasta ahora de esta influencia. En primer lugar, influir y alterar la capacidad de los ciudadanos para interpretar la realidad es un esfuerzo a largo plazo, y la IA generativa lleva disponible poco tiempo de forma amplia. En segundo lugar, la disrupción del ambiente informativo no es solo una cuestión de contenido. De hecho, tal vez el contenido en sí mismo sea el factor menos importante. Más importantes son la capacidad de difundir desinformación, los mecanismos de amplificación, la voluntad de los actores políticos de construir una realidad alternativa que no se base en hechos y, sobre todo, el fomento de grupos de personas dispuestas a abrazar esta realidad alternativa. Como se detallará más adelante, la distribución más amplia de Grok (a través de la plataforma de redes sociales X de Elon Musk) en Brasil ya ha comenzado a llenar de imágenes artificiales la política digital brasileña.

También debemos aceptar que es muy poco probable que seamos capaces de aislar de manera tajante el efecto de la IA en la integridad de la información. La forma en cómo se crea, comparte, difunde y consume la información es un proceso complejo. La IA puede influir en muchas fases diferentes de este proceso, alterándolas ligeramente y, por ello, modificando el resultado final de las elecciones. Sin embargo, en la mayoría de los casos, sería casi imposible establecer vínculos causales claros entre el uso de la IA y los cambios resultantes.

También debemos aceptar que es muy poco probable que seamos capaces de aislar de manera tajante el efecto de la IA en la integridad de la información. La forma en cómo se crea, comparte, difunde y consume la información es un proceso complejo.

A pesar de ello, la IA no hará más que mejorar con el tiempo, al igual que debería hacerlo nuestra defensa de la democracia y las elecciones. Incluso si la influencia de la IA parece marginal hasta ahora, es solo cuestión de tiempo que se vuelva más significativa. La construcción de una realidad compartida y que se base en hechos debe ser fundamental para cualquier estrategia de protección de las elecciones. Las fuerzas democráticas deben entender

que ciertos principios no son negociables, tales como: la aceptación de los resultados electorales, la garantía de la transferencia pacífica del poder, el respeto y la protección de las instituciones democráticas y la defensa del Estado de derecho. Fortaleciendo estos cimientos es como se defiende la democracia.

Este documento ahondará en el impacto potencial que la IA podría tener en la integridad de la información durante las elecciones, centrándose en tres aspectos diferentes. En primer lugar, se examinará la forma en cómo la IA impacta la creación y distribución de información. En el capítulo 1 se analizará cómo la IA generativa puede influir en la integridad de la información, explorando cómo está determinando el contenido al que tenemos acceso a través de la moderación, la curación y la distribución de dicho contenido, y las posibles consecuencias para grupos y colectivos marginados. El capítulo 2 ahondará en los casos de México y Brasil con un análisis detallado de los usos de la IA en las elecciones de estos países en 2024. En el capítulo 3 se explorarán las diferentes estrategias globales y regionales, analizando en qué medida toman en consideración e incluyen las elecciones y la democracia, al tiempo que se identifican las principales brechas. En el capítulo final se presentará un conjunto de recomendaciones de políticas para abordar los desafíos identificados.

Capítulo 1

LA IA Y LA INTEGRIDAD DE LA INFORMACIÓN EN LAS ELECCIONES

El impacto de la IA en la integridad de la información ha sido ampliamente discutido en los últimos meses, especialmente por la coincidencia, en 2024, de una serie de elecciones importantes con un incremento aparentemente exponencial de la capacidad de las herramientas de IA generativa. Sin embargo, la mayoría de los temores de que una avalancha de información poco fiable inundara el entorno informativo y comprometiera la calidad de las elecciones parecen haber sido infundados, con un impacto observable mínimo en los resultados electorales.

La aparente falta de impacto podría explicarse por una combinación de factores, como la novedad de la IA, el hecho de que aún esté en desarrollo o los esfuerzos coordinados de la sociedad civil y algunas empresas tecnológicas para evitar su uso indebido. Un buen ejemplo de este esfuerzo coordinado es el Acuerdo Tecnológico para Combatir el Uso Engañoso de la IA en las Elecciones de 2024 [*Tech Accord to Combat Deceptive Use of AI in 2024 Elections*]: una iniciativa de algunas de las principales empresas tecnológicas con sede en Estados Unidos para limitar la influencia de la IA en las elecciones.

Se esperaba que la capacidad de la IA generativa para crear contenido manipulador tuviera una gran influencia en las elecciones, en parte al arrollar el ecosistema informativo con contenido manipulador que ahogase la información fiable.

Sin embargo, el aspecto clave de los contenidos manipuladores nunca fue su oferta, sino su demanda (Simon, Altay y Mercier, 2023). Hay dos razones por las que el tener la capacidad de crear una cantidad masiva de contenido manipulador no se traduce directamente en un aumento de la capacidad de manipular a los votantes. En primer lugar, la influencia es muy difícil de lograr, y las investigaciones apuntan a la existencia de un nivel mínimo de capacidad de información en línea para que se pueda modificar el comportamiento de los votantes (Coppock, Hill y Vavreck, 2020). Los contenidos en línea creados

con el objetivo de manipular la creación de sentido colectivo pueden, a largo plazo, alterar las prioridades de la gente, pero su influencia depende menos de la cantidad de contenido que de su alcance (Rid, 2020). En segundo lugar, la demanda de contenido manipulador está altamente correlacionada con el bajo nivel de confianza en las instituciones y el partidismo fuerte (Simon, Altay y Mercier, 2023). Esto implica que el contenido manipulador se utiliza para confirmar los sesgos y las opiniones partidistas ya existentes, en lugar de transformar la percepción de los votantes.

Aunque la influencia actual de la IA sobre las elecciones pueda parecer limitada, esta percepción requiere matices y una visión a largo plazo.

Aunque la influencia actual de la IA sobre las elecciones pueda parecer limitada, esta percepción requiere matices y una visión a largo plazo. En primer lugar, la IA ya es una parte fundamental del ecosistema informativo y, como tal, influye en la forma en que se distribuye y amplifica la información. Las herramientas de IA se utilizan ampliamente para curar y jerarquizar contenidos, así como para moderarlos en todas las plataformas de redes sociales. Nuestros *feeds* están impulsados por IA, por lo que la información a la que tenemos acceso está determinada en gran medida por la IA. En segundo lugar, el perfeccionamiento exponencial de la tecnología en los últimos años apunta a que nuestra capacidad para detectarla, que ahora ya es baja, siga disminuyendo. Con aplicaciones de IA generativa mejores, más rápidas y precisas, los riesgos para la integridad de la información y la calidad de las elecciones aumentarán. En tercer lugar, la IA ya está mostrando un impacto sobre los colectivos tradicionalmente marginados, y especialmente sobre las mujeres políticas. Esto es cierto a nivel global y también para América Latina. En cuarto lugar, el análisis empírico del uso temprano de la IA generativa en las elecciones de todo el mundo muestra una diversidad de aplicaciones de la tecnología más allá de la manipulación política directa, como la producción de contenido político que apela al humor y las emociones.

Las siguientes secciones profundizarán en estos cuatro elementos, presentando un marco para la comprensión de las diferentes formas en que la IA podría afectar a la integridad de la información durante las elecciones a nivel global y en América Latina en un futuro cercano.

1.1. LA IA Y LA CREACIÓN Y DISTRIBUCIÓN DE INFORMACIÓN EN AMÉRICA LATINA

La inteligencia artificial juega un papel fundamental en la distribución de contenidos electorales en línea. Aunque herramientas de IA generativa como ChatGPT, Gemini, Microsoft Copilot y otras son populares y conocidas, las empresas tecnológicas, en particular las plataformas de redes sociales, llevan utilizando la IA mucho tiempo. La IA es una herramienta importante para distribuir y difundir contenido en las plataformas de redes sociales a través de la curación y moderación automatizadas, incluso en América Latina. Los partidos políticos, los actores estatales, los políticos y las organizaciones de la sociedad civil utilizan las plataformas de redes sociales —y, por lo tanto, los

sistemas de IA— para crear y difundir en línea contenido político relacionado con las elecciones.

Los procesos de diseño, desarrollo y despliegue de contenidos en las redes sociales no son neutrales. Por el contrario, están impulsados por incentivos políticos y económicos que dan forma a la esfera pública en línea, al entorno informativo y, en última instancia, a los procesos democráticos. Además, el marco legal institucional y el contexto geopolítico y sociopolítico también configuran el entorno de la información electoral. En consecuencia, el enfoque de las políticas conducentes a garantizar la integridad de la información electoral debe ser integral, teniendo en cuenta consideraciones políticas, culturales, sociológicas y geopolíticas, en lugar de centrarse únicamente en los ecosistemas de medios de comunicación o las plataformas de redes sociales. Este enfoque podría contribuir a iniciar un debate para la mejora de los valores democráticos en contextos de información electoral en línea.

Teniendo esto en cuenta, es importante destacar el impacto de los sistemas de IA en la integridad de la información en el contexto de los procesos electorales en América Latina. Por ello, en la siguiente sección se explicará cómo los sistemas de IA afectan a la integridad de la información en cuatro niveles: (a) creación de contenido; (b) curación de contenidos; (c) moderación de contenidos; y d) publicidad política en línea.

1.1.1. **Creación de contenido**

Las herramientas de IA generativa pueden producir contenido sofisticado a gran escala; contenido a menudo indistinguible del generado por humanos, y que atesora un potencial sin precedentes para el engaño (Juneja, 2024). En particular, la IA puede utilizarse para producir contenidos adaptados a los puntos de vista y opiniones de subconjuntos específicos de la población (Juneja, 2024). Contenidos tales como textos, elementos visuales o audios ya se han utilizado en campañas electorales en América Latina; en Argentina y Brasil, por ejemplo.

Las formas en que la IA generativa, especialmente los modelos extensos de lenguaje (LLM), pueden aumentar el alcance, el impacto y la influencia de las operaciones de información son abundantes. Es importante señalar, sin embargo, que aún no hemos sido testigos de todos los posibles efectos negativos de la tecnología en la integridad de la información durante las elecciones. Es probable que la IA generativa en general, y los LLM en particular, ejerzan principalmente un efecto multiplicador y amplificador. Este efecto tiene cuatro aspectos: a) reducción de las barreras de entrada; (b) aumento exponencial de la cantidad de contenido engañoso; c) hacer que los contenidos sean más adaptables y específicos; y (d) hacer que las operaciones de información sean más difíciles de detectar (Rid, 2020; Goldstein et al., 2023; Simon, Altay y Mercier, 2023).

La IA generativa reduce significativamente el coste y el esfuerzo para la realización de operaciones informativas, reduciendo las barreras de entrada y haciendo más accesibles dichas operaciones. Las herramientas de IA pueden

incrementar en gran medida la eficiencia de las campañas políticas. Además, pueden facilitar la expansión de los proveedores de desinformación como servicio (DaaS), permitiendo que un mayor número de actores ingresen al mercado de DaaS (Soliman y Rinta-Kahila, 2024). Sin embargo, la capacidad de emplear tales tácticas no se traduce automáticamente en la intención de hacerlo. El aprovechamiento del material manipulador para influir en la opinión pública a través del engaño y las falsedades es una estrategia adoptada solo por ciertos partidos y candidatos (Törnberg y Chueri, 2025). Para muchos grupos políticos, el empleo de prácticas engañosas no es un enfoque atractivo. La mayoría de los partidos y candidatos que adoptan estas tácticas, ya sea directamente o a través de operadores DaaS, explotan la creciente desconfianza en el sistema político y la información oficial. Su objetivo es aprovechar esta disminución de la confianza en las instituciones democráticas para obtener ventajas políticas (Törnberg y Chueri, 2025).

La IA generativa permite a los partidos políticos y candidatos dispuestos a emplear tácticas manipuladoras producir material engañoso a una escala mucho mayor. Por ejemplo, los LLM pueden facilitar la creación de contenido falso para medios de comunicación en línea fabricados, una estrategia comúnmente utilizada en operaciones de información, incluso por parte de actores extranjeros (Rid, 2024). Un aumento en el contenido engañoso puede desvincular de la realidad a segmentos de la población, con respecto a temas críticos para el futuro político de América Latina tales como el cambio climático, la corrupción política o la migración (DiResta 2024). Además, puede hacer que la gente le sea cada vez más difícil acceder a información basada en hechos, necesaria para comprender el mundo con precisión.

De cualquier manera, el panorama de la información digital ya está inundado de contenido engañoso, lo que sugiere que solo un aumento de la demanda, en lugar de la oferta, convertiría la mayor capacidad de generar dicho material en una verdadera amenaza para la democracia (Simon, Altay y Mercier, 2023). Además, inundar el espacio informativo no equivale necesariamente a influencia. La desinformación afecta principalmente a aquellas personas que ya se inclinaban a creerla, en tanto refuerza sus percepciones preexistentes de la realidad (DiResta 2024). Factores como la desconfianza en las instituciones democráticas y la fuerte identificación partidista, sobre todo con los partidos de derecha radical, hacen que la gente sea más susceptibles a la adopción de relatos falsos (Grinberg et al., 2019; Törnberg y Chueri, 2025).

La IA generativa reduce significativamente el coste y el esfuerzo para la realización de operaciones informativas, reduciendo las barreras de entrada y haciendo más accesibles dichas operaciones.

Un tercer impacto potencial de la IA generativa en la integridad de la información es su capacidad para la personalización extrema del contenido engañoso. Dichos contenidos pueden utilizarse para dañar a la gente, sobre todo mediante la creación de imágenes no consentidas, o para manipular sus percepciones. Una amenaza significativa planteada por la IA es el contenido pornográfico no consentido (para un análisis en profundidad, véase 1.2: La IA y las comunidades marginadas). Este tipo de contenido afecta de manera desproporcionada a las figuras políticas femeninas, actuando como una forma de violencia de género facilitada por la tecnología que limita aún más el espacio político disponible para las mujeres.

La IA generativa también aumenta la capacidad de adaptar y dirigir el contenido a grupos sociales específicos, lo que en teoría incrementa el poder persuasivo del contenido engañoso. Entre los métodos para adaptar y orientar el contenido está el uso de un lenguaje culturalmente apropiado, el empleo de referencias específicas, el encuadre de los problemas de una manera que cale en el grupo objetivo o la elaboración de contenido que manipule la realidad para atraer a grupos demográficos específicos. Si bien muchas operaciones de información dirigidas por extranjeros pueden detectarse debido a su uso antinatural del lenguaje o a la falta de conocimiento cultural, la IA generativa puede sortear estos escollos (Goldstein et al., 2023).

Incluso es técnicamente concebible que, en el futuro, las operaciones de información se vuelvan mucho más granulares, dirigiéndose a individuos con contenido engañoso adaptado a sus perfiles. Estas operaciones podrían aprovechar datos individualizados para generar materiales manipuladores, diseñados para influir o perjudicar a personas específicas. Hasta la fecha, sin embargo, los modelos de IA generativa no tienen acceso a datos personalizados que permitan este tipo de contenido.

En el caso de América Latina, la tecnología, en su forma actual, podría tener dificultades para dirigirse a la gente a un nivel tan granular. Los modelos de IA se entrenan a menudo con contenido creado originalmente en inglés y luego traducido automáticamente al español. En consecuencia, por un lado, los estudios han sugerido que estos modelos perpetúan prejuicios raciales, lo que podría impactar negativamente en el poder de persuasión del contenido (Hoffmann et al., 2024). Dado que están diseñados en inglés y funcionan mejor en contextos de habla inglesa, generan un desajuste entre los angloparlantes nativos y los usuarios de otros lugares, incluyendo América Latina (Nicholas y Bhatia, 2023), donde la falta de contexto local y datos en otros idiomas plantea un problema sistemático en la forma en que se utilizan estas tecnologías (Oversight Board, 2024).

Por último, la IA generativa podría mejorar la capacidad de las operaciones de información, sobre todo las realizadas por actores extranjeros, para evadir la detección (Goldstein et al., 2023). Las operaciones de influencia llevadas a cabo por potencias extranjeras fracasan a menudo debido a la falta de precisión en sus mensajes y el uso del lenguaje. Además, las empresas de redes sociales y algunas agencias estatales suelen tener la capacidad de identificar comportamientos inauténticos coordinados —típicos de este tipo de operaciones— y de detectar estas acciones antes de que puedan tener un efecto real (Rid, 2024). Sin embargo, la IA generativa puede dificultar la detección de operaciones de información mediante la creación de redes de identidades falsas en línea distribuidas de manera que se oculte su origen. Estas identidades pueden utilizarse para propagar discursos y forjar realidades alternativas, ya que pueden imitar mejor los dialectos locales, y el uso del idioma o incluso fingir acentos. También pueden saturar a las organizaciones de verificación de datos y dificultar grandemente su trabajo (Kahn, 2024).

La IA generativa puede dificultar la detección de operaciones de información mediante la creación de redes de personas falsas en línea que se distribuyen de una manera que oculta su origen.

1.1.2. **Curación de contenidos**

La IA se está utilizando, no solo para crear contenido en las plataformas de redes sociales, sino también para jerarquizar, promocionar o relegar contenido en las mismas, generalmente en función no solo de las preferencias y perfiles de los usuarios, sino también del uso de contenido diseñado para mantenerlos lo más enganchados posible, sin tener en cuenta ni su calidad ni su precisión (ARTICLE 19, 2021; OSCE, 2021; Narayanan, 2023). El objetivo principal de las herramientas impulsadas por IA, a menudo denominadas "algoritmos de clasificación" [*ranking algorithms*], es ayudar a las personas a encontrar contenido relevante en función de las consideraciones comerciales y, en algunos casos, políticas de las plataformas que utilizan. Dado el volumen, la escala y la velocidad a la que se propaga el contenido en las redes sociales, los sistemas de IA desempeñan un papel decisivo a la hora de priorizar y degradar el contenido en los *feeds* de los usuarios en función de sus perfiles y los modelos de negocio de las plataformas.

Estos avances pueden afectar a la precisión, la diversidad y el interés público del ecosistema informativo en los espacios en línea, lo que es especialmente sensible durante las elecciones (OSCE, 2021). Naciones Unidas (2018) también ha dicho que los usuarios pueden tener poca o ninguna exposición a ciertos tipos de contenido social o político crítico en sus plataformas. Esta falta de exposición podría socavar no solo la libertad de los ciudadanos de acceder a información diversa, sino también su derecho a formarse opiniones de manera independiente, libre de violencia, inducción o interferencia manipuladora de cualquier tipo (ACNUDH, 1996). Tales limitaciones al acceso a diversos contenidos podrían implicar una violación de otros derechos fundamentales, incluido el derecho a participar en los asuntos públicos.

En consecuencia, las herramientas impulsadas por IA que recomiendan contenido no son neutrales. La curación algorítmica viene determinada por los valores y objetivos del creador del algoritmo, así como por la dinámica sociotécnica, la autorregulación de la industria y la supervisión estatal (OSCE, 2021). Estos algoritmos se alimentan de un gran volumen de datos personales y de comportamiento, que se combinan con datos de terceros para crear perfiles sofisticados orientados a aumentar la participación y la retención de los usuarios, mediante la personalización de su experiencia.

La personalización del contenido también se utiliza en la publicidad y en las campañas políticas en línea. Se ha señalado que esta práctica puede implicar manipulación, polarización y fragmentación del debate político (Gorton, 2016). Además, varios estudios han demostrado que los algoritmos de curación o clasificación de contenidos contribuyen a la amplificación de la desinformación, el discurso de odio, los contenidos nocivos y la discriminación en las plataformas de redes sociales (Fundacja Panoptykon, 2023). El "contenido pronosticado más atractivo obtiene las puntuaciones más bajas en términos de calidad y confiabilidad" (Fundacja Panoptykon, 2023, pág. 3). La Fundación Mozilla (2021) también informó en un estudio que el 71 por ciento de las experiencias desagradables en YouTube planteadas por voluntarios habían sido recomendadas a los espectadores por la plataforma. Dado que

es más probable que las personas hagan clic en contenido sensacionalista o incendiario, los algoritmos de clasificación privilegian el contenido dañino (Amnistía Internacional, 2019) en lugar de la información precisa, lo que socava la integridad electoral tanto en contextos *online* como *offline*.

En Brasil, durante el régimen de Bolsonaro—especialmente hasta las elecciones de 2022—se difundieron campañas de desinformación y teorías conspirativas en Twitter y WhatsApp, con el objetivo de desacreditar el voto electrónico y el proceso democrático del país (Souza, 2024). El proceso de radicalización iniciado en 2018, con la primera campaña presidencial de Jair Bolsonaro, y que culminó con el intento de golpe de Estado contra el presidente electo, Luiz Inácio Lula da Silva en 2022, muestra el impacto que puede tener la desinformación.

Por otro lado, las herramientas de curación de contenidos impulsadas por IA se entrenan con datos que no reflejan realidades contextuales, históricas y económicas más amplias. Por lo tanto, es posible que aparezcan sesgos en la forma en que se distribuye el contenido, distorsionando, minimizando o amplificando ciertos mensajes para involucrar a los ciudadanos en las plataformas de redes sociales.

Debido al volumen de datos procesados por las herramientas impulsadas por IA, los algoritmos de clasificación deben examinarse en dos niveles. En primer lugar, los datos que se recopilan para entrenarlos deben divulgarse y categorizarse claramente. En segundo lugar, los algoritmos de clasificación deben explicar el razonamiento que hay detrás de su producción (por qué se muestra cierto contenido a un usuario en particular) y qué puntos de datos llevaron a la decisión de mostrar ese contenido.

El diseño de incentivos que promuevan los valores democráticos, como la pluralidad de los medios de comunicación así como la diversidad de opiniones, y la restricción de los datos utilizados para entrenar los algoritmos, podrían mitigar los efectos negativos de la curación de contenidos. Proporcionar acceso a los datos y permitir auditorías de algoritmos por parte de terceros podría ayudar a reducir los resultados inesperados y problemáticos de estas herramientas impulsadas por IA.

El diseño de incentivos que promuevan los valores democráticos, como la pluralidad de los medios de comunicación y la diversidad de opiniones, así como la restricción de los datos utilizados para entrenar los algoritmos, podrían mitigar los efectos negativos de la curación de contenidos.

1.1.3. **Moderación de contenidos**

La moderación de contenido implica a menudo la externalización a gran escala combinada con la detección automatizada de aprendizaje automático [*machine learning*]. Los sistemas de IA ayudan a las empresas de redes sociales a identificar, combinar, predecir o clasificar el contenido generado por el usuario en función de sus características para determinar qué debe permanecer en línea y qué debe eliminarse. Este proceso suele darse después de que los usuarios publican contenido en una plataforma de redes sociales. En otras palabras, las herramientas de moderación de contenidos actúan como guardianes, guiándose por las políticas y el modelo de negocio de la plataforma, así como por el marco jurídico nacional en el que la plataforma opera.

Las herramientas de moderación de contenido pueden desempeñar las siguientes funciones:

- detectar contenidos que infrinjan la ley (por ejemplo, incitación al odio o difusión de imágenes íntimas sin consentimiento) o las políticas internas de la plataforma;
- identificar las infracciones evaluando si el contenido detectado infringe las políticas internas; y
- hacer cumplir las políticas marcando o eliminando el contenido en línea creado por los usuarios, en función de las directrices de la plataforma (OSCE, 2021).

Las plataformas emplean a menudo prácticas como la despriorización y la desjerarquización para minimizar la visibilidad de contenido dañino o ilegal.[1] La suspensión y eliminación de cuentas también es parte integral de estos procesos de moderación de contenido.

Los sistemas de IA facilitan estos procesos al revisar, tomar decisiones y restringir la información o la desinformación a gran escala, a menudo sin intervención humana. La automatización de estos procesos puede mejorar la coherencia y la previsibilidad. Gracias al procesamiento del lenguaje natural, las técnicas de moderación de contenidos pueden identificar emociones positivas y negativas, clasificándolas como "ofensivas" o "no ofensivas".

Sin embargo, el uso de la IA en la moderación de contenidos también comporta riesgos. El etiquetado incorrecto puede dar lugar tanto a falsos positivos, con contenido legítimo etiquetado erróneamente como ilegal o perjudicial, como a falsos negativos, con contenido ilegal o perjudicial pasando desapercibido.

Los desafíos antes mencionados se ven agravados por los sesgos en los datos de entrenamiento utilizados para los sistemas de IA, que a menudo se basan en contenido escrito principalmente en inglés. Esta dependencia del contenido en inglés crea dificultades para detectar la desinformación o el discurso de odio en contextos no ingleses. Por ejemplo, para el caso del quechua, una lengua indígena hablada en América Latina, se ha informado de que el “mayor obstáculo para abordar los daños en línea es la ‘falta de datos digitales de alta calidad’” (Shahid, 2024). Además, los “Facebook Papers” revelaron que el 87 por ciento del gasto global de Facebook en clasificar desinformación se dirigió al contenido en inglés, a pesar de que solo el 9 por ciento de sus usuarios son hablantes nativos de ese idioma (Elswah, 2024).

[1] Los procesos señalados es lo que en inglés se ha denominado “*shadow banning*” o “Bloque en la sombra”, en donde las redes sociales o plataformas restringen disimuladamente contenido, con el fin de ocultarlo del publico sin removerlo totalmente de la plataforma.

1.1.4. **Publicidad política en línea**

Las plataformas en línea dependen de la publicidad en línea, incluida la publicidad política *online*, para monetizar sus servicios. La publicidad está en el centro de los modelos de negocio de las plataformas de redes sociales porque les permite monetizar la atención y la retención de los usuarios. En este contexto, los sistemas de IA se utilizan para desplegar y distribuir publicidad política en las plataformas de redes sociales.

Por un lado, las técnicas de microsegmentación (segmentación de anuncios) permiten a partidos políticos y activistas identificar y dirigirse a segmentos específicos del electorado con anuncios políticos personalizados de alta granularidad. Es posible, por ejemplo, presentar un anuncio político particular a mujeres jóvenes que viven en Santiago de Chile, de 18 a 24 años, a quienes les gusta la música K-pop, se inclinan políticamente hacia la izquierda y se interesan por la naturaleza y los deportes. Un anuncio distinto, con un mensaje completamente diferente, podría mostrarse a otro grupo demográfico específico. Estas técnicas de segmentación se basan en diferentes fuentes de datos, incluidos los datos personales y de comportamiento, los datos generados por las propias plataformas de redes sociales y los datos de terceros. En conjunto, estos datos se utilizan para crear perfiles altamente personalizados que determinan qué contenido se amplifica en el *timeline* de un usuario en particular y qué anuncios se muestran al usuario (Ali et al., 2019).

Por otro lado, las plataformas de redes sociales utilizan técnicas de amplificación para entregar anuncios políticos a los destinatarios más relevantes (European Parntership for Democracy, s.f.). Este proceso, conocido como *ad delivery* [entrega de anuncios], implica que la plataforma muestre anuncios a los usuarios en función de algoritmos de aprendizaje automático. Estos algoritmos procesan datos de comportamiento como los *likes*, los *posts* compartidos, las conexiones, los clics y las visualizaciones de video para crear perfiles completos de usuario en línea. Con estos datos, los sistemas de IA pueden predecir los comportamientos de los usuarios, revelando características sensibles como la inclinación política, las creencias religiosas, las condiciones de salud, etcétera. Sin embargo, debido a que algunos usuarios son considerados más valiosos que otros, este proceso puede tener consecuencias discriminatorias en la distribución de anuncios políticos (Ali et al., 2019).

Además, estas técnicas ayudan a afinar los algoritmos de clasificación para mantener a los usuarios enganchados al contenido, lo que puede reforzar los sesgos y alimentar las opiniones radicales. Al difundir desinformación y mala información, contribuyen a la creación de cámaras de eco, que pueden fragmentar la esfera pública en línea. Esta fragmentación facilita la manipulación, la polarización y la radicalización de la ciudadanía.

1.2. LA IA Y LOS GRUPOS MARGINALIZADOS

Al igual que ocurre con la mayoría de los desafíos sociales, los grupos marginados sienten el impacto de esta cuestión de manera más aguda que el resto de la población. En América Latina, eso incluye colectivos de clase trabajadora, mujeres, personas LGBTQIA+, personas con discapacidad y grupos indígenas. Los grupos marginados se enfrentan a riesgos cada vez mayores, incluso en los espacios digitales, que pueden verse agravados por sistemas de IA diseñados, implementados o regulados sin la debida consideración de los derechos humanos y los valores democráticos.

Uno de los ejemplos más claros de las consecuencias potencialmente negativas de la IA es su capacidad para perpetuar y amplificar la discriminación y el acoso por motivos de género en los entornos en línea.

Uno de los ejemplos más claros de las consecuencias potencialmente negativas de la IA es su capacidad para perpetuar y amplificar la discriminación y el acoso por motivos de género en los entornos en línea. La IA generativa, por ejemplo, se ha utilizado como arma en todo el mundo para crear contenido *deepfake,* incluido material explícito no consensuado, que se dirige de manera desproporcionada a las mujeres y las minorías sexuales (Burgess, 2021). Numerosos sitios *web* ofrecen servicios de *deepfake,* lo que supone una amenaza tangible para las mujeres, que a menudo son objeto de contenidos pornográficos generados por IA. Las aplicaciones permiten a los usuarios subir fotos y generar imágenes de desnudos realistas de personas sin su consentimiento. En México, por ejemplo, un estudiante universitario transformó 166.000 fotos de compañeros de clase en material pornográfico y las vendió en Telegram (Rodríguez, 2023).

Estas prácticas no se limitan a la explotación personal, sino que también se utilizan indebidamente para debilitar a las candidatas políticas. En las recientes elecciones municipales de 2024 en Brasil, por ejemplo, se empleó tecnología *deepfake* para dirigirse a candidatas políticas, representando a mujeres en escenarios comprometedores o explícitos, con el objetivo de dañar su reputación y sus perspectivas electorales (Farrugia, 2024). Si bien los políticos masculinos también son blanco de los *deepfakes,* a menudo por sus declaraciones o acciones políticas, estos ataques les permiten generalmente preservar un sentido de integridad personal. Por el contrario, los ataques *deepfake* contra mujeres políticas tienden a girar en torno a conductas personales indebidas, que alimentan la cosificación, refuerzan estereotipos dañinos y minan la credibilidad de dichas mujeres (Cerdán Martínez y Padilla Castillo, 2019). El uso lesivo de la tecnología *deepfake* no solo compromete la dignidad de las mujeres, sino que también puede disuadirlas de presentarse a las elecciones o de participar plenamente en el proceso electoral (Souza y Varon, 2019; Fundación Multitudes, 2023; Ríos Tobar, 2024), con repercusiones sobre su bienestar mental, seguridad personal y representación política en general.

Varios países latinoamericanos han comenzado a adoptar enfoques de género en las políticas de ciberseguridad, abordando problemas como el ciberacoso y el intercambio no consentido de imágenes íntimas (González-Véliz y Cuzcano-Chávez, 2024). Las elecciones municipales de Brasil de octubre de 2024 se celebraron bajo una nueva y estricta normativa electoral que prohibía el uso de

contenidos generados por IA no etiquetados en las campañas, en un esfuerzo por frenar la desinformación impulsada por estas tecnologías. Aunque es probable que estas medidas reduzcan algunos riesgos, la regulación por sí sola no es suficiente para prevenir por completo la propagación de *deepfakes* dañinos y violentos. Del mismo modo, la Ley Olimpia de México de 2020 penaliza la difusión no consentida de contenidos sexuales, aunque su alcance sigue siendo demasiado limitado para abordar plenamente los delitos específicos de la tecnología *deepfake* (Piña, 2023).

Además, las plataformas de redes sociales que dependen de la IA para la moderación de contenidos suelen quedarse cortas a la hora de proteger a los grupos marginados. Por ejemplo, las investigaciones muestran que la moderación de contenidos impulsada por la IA en las plataformas de redes sociales suele tener dificultades para identificar y eliminar eficazmente el discurso de odio dirigido contra grupos marginados, a menudo debido a la naturaleza matizada y específica del contexto de dichos contenidos (Udupa, Maronikolakis, Schütze y Wisiorek, 2023). Por el contrario, los sistemas de IA señalan a veces el activismo o los debates legítimos sobre cuestiones de género y LGBTQIA+ como violaciones de las políticas de las plataformas, silenciando las voces críticas y limitando su alcance.

Las comunidades indígenas suelen ser particularmente vulnerables a la manipulación de los entornos de información electoral. Se enfrentan a una serie de desafíos, como las campañas de desinformación, la manipulación e injerencia de la información extranjera, la exclusión sistemática de los espacios digitales y la amplificación de los discursos de odio por medio de las tecnologías emergentes. Con frecuencia, estas campañas explotan estereotipos culturales o tergiversan las tradiciones y demandas políticas indígenas, lo que socava la credibilidad de los líderes, activistas y movimientos indígenas (Aaberg, Gurung, Kandie, Méndez y Xón, 2024).

Durante los períodos electorales en Bolivia, por ejemplo, la desinformación, el discurso de odio y los discursos cargados de racismo han circulado deliberadamente para deslegitimar la participación indígena en la política (Wood, 2019). Tales tácticas no solo distorsionan el discurso público, sino que también incitan al odio, el racismo y los prejuicios contra los pueblos indígenas, erosionando efectivamente su importancia y contribuciones tanto a nivel nacional como mundial. Los informes han documentado el uso de información engañosa para fomentar la intolerancia, profundizando así los agravios históricos y las injusticias sistemáticas contra las comunidades indígenas (The Carter Center, 2021; Funk, Shahbaz y Vesteinsson, s.f.).

Las tecnologías de IA han exacerbado la vulnerabilidad de las comunidades indígenas. Las herramientas automatizadas de generación de contenidos, las redes de *bots* a gran escala y la segmentación algorítmica sofisticada pueden incrementar en gran medida la propagación, la personalización y la persistencia de la desinformación y el discurso de odio (Funk, Shahbaz y Vesteinsson, s.f.). Como resultado de ello, los líderes y movimientos indígenas pueden encontrarse con una desventaja aún mayor, luchando por contrarrestar

los discursos que se difunden amplia y persistentemente a través de las plataformas digitales.

Los sesgos algorítmicos intensifican las formas existentes de exclusión, especialmente cuando los sistemas de IA se entrenan con conjuntos de datos que no representan adecuadamente a las comunidades indígenas. Como resultado, estas tecnologías refuerzan estereotipos dañinos y no satisfacen las necesidades específicas de las poblaciones indígenas. Las deficiencias metodológicas, como la generalización excesiva o la falta de adaptación de los modelos a los contextos locales, pueden distorsionar la realidad. Al hacerlo, los algoritmos sesgados no solo tergiversan la verdad, sino que también amplifican las desigualdades existentes, ensanchando así la brecha digital a la que se enfrentan las comunidades indígenas (Naciones Unidas en México, 2023).

Antes de plantearse el uso de la IA en procesos públicos como las elecciones, hay que abordar la cuestión fundamental del acceso. En toda América Latina y el Caribe, las comunidades indígenas se enfrentan a una brecha digital étnica pronunciada, ya que la infraestructura y la conectividad atienden principalmente a las zonas urbanas. Este desequilibrio deja a muchos pueblos indígenas de las regiones rurales con oportunidades limitadas para acceder a las herramientas digitales. Para agravar este desafío, las brechas generalizadas en la alfabetización digital restringen aún más su capacidad para participar plenamente en el ecosistema digital. Según la Comisión Económica para América Latina y el Caribe de las Naciones Unidas, mientras que el 40 por ciento de la población de la región tiene conocimientos básicos de informática, menos del 30 por ciento puede usar hojas de cálculo de manera efectiva, menos del 25 por ciento puede instalar nuevos dispositivos o *software* y solo alrededor del 7 por ciento tiene experiencia en programación (Naciones Unidas en México, 2023).

Construir una IA ética e inclusiva exige que cerremos las brechas de conectividad y habilidades, corrijamos los sesgos algorítmicos, defendamos la soberanía de los datos indígenas e integremos diversas visiones del mundo.

Estas desigualdades sistemáticas perpetúan la exclusión de las perspectivas indígenas del desarrollo de la IA. La conectividad limitada, las políticas insuficientes y las competencias técnicas inadecuadas refuerzan un ciclo de marginación que disminuye el conocimiento y las experiencias indígenas. Construir una IA ética e inclusiva exige que cerremos las brechas de conectividad y habilidades, corrijamos los sesgos algorítmicos, defendamos la soberanía de los datos indígenas e integremos diversas visiones del mundo. Un primer paso crucial es reconocer a los pueblos indígenas no solo como participantes, sino también como creadores e innovadores dentro de los marcos tecnológicos.

Estos son solo algunos de los desafíos que ilustran la intrincada superposición entre la integridad de la información, la IA y la integridad electoral, en particular su impacto desproporcionado sobre los grupos marginados de la sociedad. Este documento busca explorar el papel de la IA en estos procesos electorales

y desarrollar recomendaciones específicas para que las empresas de redes sociales, los responsables políticos y las autoridades regulen de manera eficaz los sistemas de IA. Dichas regulaciones deben abordar la creación, la curacióny la moderación de contenidos en línea, dando prioridad a los derechos humanos y los principios democráticos.

Capítulo 2

IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN LAS ELECCIONES DE MÉXICO Y BRASIL

Para comprender cómo la inteligencia artificial ha influido en los procesos electorales en América Latina, este capítulo investigará cómo se utilizaron las herramientas impulsadas por IA en las elecciones de 2024 en México y Brasil. Para este análisis, nos basamos en informes elaborados por FGV Comunicação Rio (2024) sobre el uso de la inteligencia artificial generativa en elecciones en todo el mundo en 2024, en artículos publicados en sitios web de noticias brasileños y mexicanos y en el monitoreo continuo del debate público realizado por DappLab (FGV DAPP, 2017), el laboratorio de monitoreo de redes sociales FGV Comunicação Rio.

Las elecciones que tuvieron lugar en México el 2 de junio de 2024 estuvieron entre las más grandes de la historia del país, con 92 millones de votantes habilitados y una participación final de más del 60 por ciento (INE, 2024a). Los comicios se llevaron a cabo a través del voto directo y obligatorio, utilizando boletas de papel, participaron siete partidos políticos y más de 20.000 cargos electos, incluidos diputados, senadores, el presidente, gobernadores estatales, autoridades locales y un nuevo alcalde de Ciudad de México, la capital del país. Operando bajo un sistema presidencial republicano, el parlamento nacional es bicameral, compuesto por una cámara con 500 diputados y otra con 128 senadores.

Tres candidatos se disputaron principalmente la carrera presidencial: Claudia Sheinbaum, en representación de la coalición gobernante de izquierda Sigamos Haciendo Historia, que ganó con el 59,3 por ciento de los votos válidos; Xóchitl Gálvez, de la alianza opositora Fuerza y Corazón, que obtuvo el 27,9 por ciento de los votos válidos; y Jorge Álvarez Máynez, del Movimiento Ciudadano, que obtuvo el 10,4 por ciento de los votos válidos (INE, 2024a).

Las elecciones de 2024 marcaron un hito para la representación de las mujeres en la política mexicana, ya que se eligieron cuatro gobernadoras y la primera mujer presidenta del país (Martínez Holguín, 2024)

Como sucesora del presidente Andrés Manuel López Obrador, Sheinbaum encabezó constantemente las encuestas, enfrentándose a críticas similares a las dirigidas a López Obrador, sobre todo en lo que respecta a la supuesta indulgencia del gobierno de izquierda de México hacia los cárteles de la droga. En este contexto, la violencia y la delincuencia dominaron la preocupación pública, convirtiéndose en temas centrales de los debates políticos, particularmente en relación con la alta tasa de homicidios del país, la violencia política de género, las disputas entre cárteles y los incidentes de violencia política. A pesar de estos problemas, los temas relacionados con la economía y la salud también adquirieron protagonismo, mientras que las políticas migratorias se discutieron de manera más marginal (Associated Press, 2024).

Al mismo tiempo, la desinformación siguió siendo un problema persistente antes de las elecciones. A lo largo del período electoral circularon rumores de fraude electoral, una propuesta de nueva constitución, la abolición de la propiedad privada e incluso afirmaciones de que Sheinbaum tenía la intención de cerrar la basílica de Nuestra Señora de Guadalupe. Los informes locales también destacan la difusión de información falsa, incluidos los vínculos inventados entre López Obrador, Sheinbaum y el narcotráfico, así como las acusaciones de que la esposa de López Obrador tenía conexiones con el nazismo.

En este contexto, las elecciones mexicanas también sirvieron de telón de fondo para el uso de la IA generativa en las campañas políticas y la información, no solo para difundir *deepfakes* para desacreditar a los candidatos o difundir desinformación política, sino también para las estafas financieras que explotaban la imagen de los candidatos.

Brasil celebró elecciones municipales en octubre de 2024, en las que se eligieron alcaldes, vicealcaldes y representantes legislativos en 5.569 municipios. Si bien las elecciones municipales en el país tienen una tradición de menor polarización y reciben menos recursos que las nacionales, las elecciones de 2024 se llevaron a cabo en un ambiente político muy cargado. Las contiendas en varias ciudades clave, entre ellas São Paulo y Fortaleza, fueron particularmente intensas, con candidatos a la alcaldía alineados estrechamente con figuras políticas nacionales como el presidente Luiz Inácio Lula da Silva y el expresidente Jair Bolsonaro, quien, a pesar de no ser elegible para ocupar cargos públicos, sigue siendo una figura central de la oposición. Estas elecciones locales fueron ampliamente vistas como un preludio estratégico para las elecciones generales de 2026, forjando alianzas y discursos para la próxima carrera presidencial.

Las elecciones se celebraron en medio de una creciente preocupación por el papel de las plataformas digitales en la democracia brasileña. Dada la anticipación en torno a la posible influencia de la IA sobre las campañas políticas, la principal autoridad electoral de Brasil, el Tribunal Superior Electoral, emitió nuevas regulaciones antes de las elecciones para frenar la desinformación y el uso indebido de contenido generado por IA (TSE, 2024).

Estas medidas incluían el etiquetado obligatorio de contenido sintético en los anuncios políticos y una supervisión más estricta de las campañas digitales.

Contrariamente a lo esperado, la IA no se utilizó tan ampliamente como se temía en principio.

Contrariamente a lo esperado, la IA no se utilizó tan ampliamente como se temía en principio. Al igual que en México, gran parte del uso observado de la IA durante el periodo electoral se vio eclipsado por una serie de acontecimientos que, si bien se vinculaban al uso de Internet en la política, no estaban relacionados directamente con la inteligencia artificial. Estos eventos y los casos documentados de uso de la IA revelan una relación compleja entre la innovación y los desafíos de larga data en la comunicación política brasileña. Incluso en esta etapa temprana, las características de estas aplicaciones de IA son notables.

2.1. **USO DE LA INTELIGENCIA ARTIFICIAL GENERATIVA EN LAS ELECCIONES NACIONALES Y LOCALES**

En México, el uso de la inteligencia artificial sigue siendo considerado un tema periférico en comparación con problemas más acuciantes, como la violencia política y el narcotráfico. Aunque no ha habido muchos casos que hayan tenido repercusiones significativas para las plataformas digitales, la difusión de *deepfakes* y otras aplicaciones de IA está vinculada a problemáticas nacionales, particularmente en el contexto electoral, donde estas tecnologías se han utilizado de manera constructiva tanto como herramientas para la desinformación y la manipulación (Del Pozo y Arroyo, 2024).

En estas elecciones que dieron como resultado la primera mujer presidenta del país, la desinformación reflejó problemas políticos de larga data en México. Los usos potencialmente dañinos de la IA, especialmente los *deepfakes,* también siguieron esta tendencia, atrayendo una mayor atención dentro del panorama electoral mexicano. Ejemplos notables incluyen videos manipulados del presidente López Obrador, en los que promovía presuntamente esquemas de inversión fraudulentos, como acciones de Pemex o aplicaciones financieras falsamente atribuidas a Elon Musk. Este material incluso fue verificado por Reuters tras ser compartido más de 2.000 veces en Facebook (El Economista, 2024).

Es interesante observar que, incluso en el contexto electoral, se ha prestado mucha atención a las falsificaciones dirigidas contra los sectores económico y financiero, y no tanta a las directamente relacionadas con las elecciones. En otras palabras, el contenido fraudulento utilizaba imágenes de figuras políticas, pero no mencionaba específicamente las elecciones.

Antes de ser elegida presidenta, Sheinbaum también fue objeto de contenido manipulado que la vinculaba con esquemas de inversión fraudulentos, mientras los *deepfakes* regionales ganaban protagonismo en las disputas locales. En Tijuana, se utilizaron grabaciones de audio atribuidas a figuras políticas para influir en las campañas, y en el estado de San Luis Potosí,

Enrique Galindo Ceballos, candidato a la alcaldía, denunció el uso de la IA para lanzar ataques contra su candidatura, lo que llevó a una denuncia formal ante el Instituto Nacional Electoral (Flores Saviaga y Savage, 2024).

Figura 2.1. **Captura de pantalla de un *deepfake* de la candidata presidencial mexicana: Claudia Sheinbaum promocionando una plataforma financiera**

Fuente: Mexico News Daily, "Video circulating of Claudia Sheinbaum is seem a 'deepfake'", 26 de enero de 2024, <https://mexiconewsdaily.com/politics/video-circulating-of-claudia-sheinbaum-is-apparently-a-deepfake>, fecha de consulta: 24 de junio de 2025.

A nivel nacional, los *deepfakes* apuntaban a los principales candidatos presidenciales. Sheinbaum, por ejemplo, fue representada en videos manipulados con símbolos comunistas de fondo mientras hablaba en ruso, mientras que Gálvez apareció sosteniendo una bandera nacional con un emblema invertido (Flores Saviaga y Savage, 2024).

El uso de la IA también se observó en las elecciones locales de México, con la Fundación Friedrich Naumann y el Consejo Alemán de Relaciones Exteriores (Del Pozo y Arroyo, 2024) identificando 44 casos de este tipo. Un ejemplo destacado fue un clip de audio supuestamente proveniente de una conversación entre Montserrat Caballero, alcaldesa de Tijuana, y Maricarmen Flores, entonces candidata por el Partido Acción Nacional, discutiendo temas de campaña (Del Pozo y Arroyo, 2024).

Otros usos de la IA revelan cómo las figuras políticas pueden aprovechar la tecnología para protegerse de las acusaciones, alegando haber sido víctimas de contenido fabricado, incluso cuando tales afirmaciones seguían sin probarse. En las elecciones mexicanas, Martí Batres, candidato a la presidencia, afirmó que una grabación de audio que circulaba en redes sociales en la que discutía el proceso interno de selección de candidatos del partido Morena había sido generada por inteligencia artificial. A pesar de su

defensa, los expertos dijeron a los periódicos mexicanos en ese momento que era imposible determinar si la grabación era válida o generada por IA (Murphy, 2024).

En las elecciones brasileñas también se observaron casos similares de grabaciones de audio de autenticidad discutida, con los propios candidatos impugnando su validez. Sin embargo, muchos de estos casos no fueron confirmados o fueron desestimados por el Tribunal Electoral del país debido a la falta de pruebas o de las herramientas técnicas necesarias para verificar la autenticidad de las grabaciones (Murphy, 2024).

El caso más notable ocurrió en la segunda vuelta de las elecciones brasileñas, cuando el candidato a la alcaldía de Fortaleza, André Fernandes, acusó a su oponente Evandro Leitão y a su equipo de crear una grabación de audio supuestamente falsa en la que Fernandes parecía respaldar el soborno y la compra de votos. Fernandes hizo la acusación en Instagram, convocó una conferencia de prensa y presentó las pruebas ante el Tribunal Electoral. Al igual que en el caso de Batres, los medios de comunicación consultaron a expertos para verificar la autenticidad de la afirmación, pero no pudieron determinar si el audio era realmente falso.

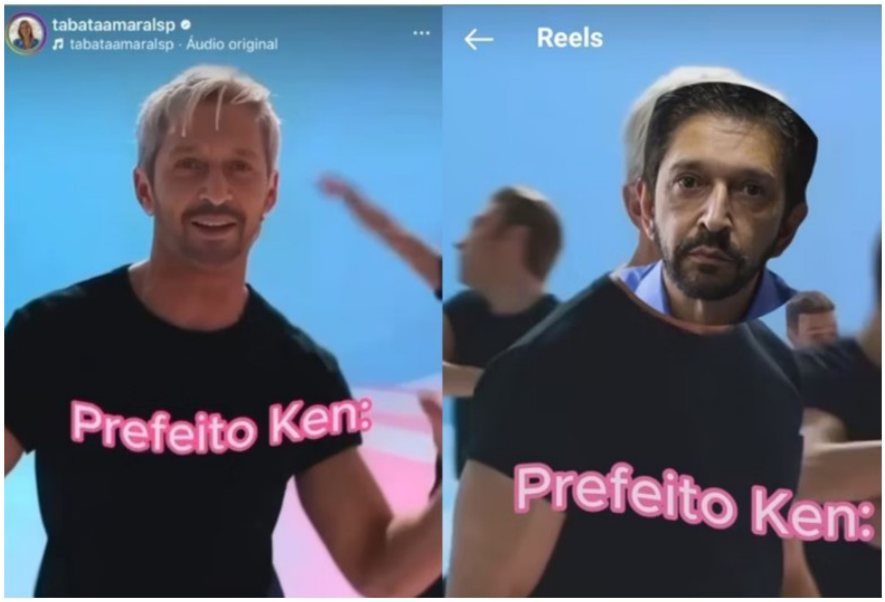
Otro uso de herramientas de IA generativa observado en las elecciones analizadas fue la difusión de *cheapfakes,* es decir, manipulaciones sencillas y de bajo coste de imágenes y vídeos, como cortes de edición, cambios de velocidad o recontextualización. A diferencia de los *deepfakes,* que utilizan IA avanzada para crear falsificaciones realistas, los *cheapfakes* se basan en ediciones básicas que el público general puede hacer.

Uno de los ejemplos más representativos de esta técnica en las elecciones brasileñas ocurrió en abril de 2024, antes del periodo oficial de campaña para las elecciones municipales. En ese momento, Tabata Amaral, precandidata a la alcaldía de São Paulo, se burló del entonces alcalde, Ricardo Nunes, después de que se refiriera a ella como la "Barbie de la política". En respuesta, Amaral publicó un video de la película de *Barbie* con la cara de Nunes superpuesta al personaje de Ken. El partido de Nunes, el Movimiento Democrático Brasileño, presentó una denuncia ante el Tribunal Electoral, alegando que el uso de la IA estaba prohibido y acusando a Amaral de hacer campaña prematura. En respuesta, la diputada eliminó el video y publicó una nueva versión, cambiando de una edición de estilo *deepfake* a una estética de collage más *amateur* (Resk, 2024).

Otros usos de la IA revelan cómo las figuras políticas pueden aprovechar la tecnología para protegerse de las acusaciones, alegando haber sido víctimas de contenido fabricado, incluso cuando tales afirmaciones seguían sin probarse.

Se observaron ejemplos similares durante las elecciones, en videos compartidos tanto por los propios candidatos como por sus partidarios. En estos casos, los candidatos y otras figuras políticas prominentes se veían colocados en situaciones inusuales y vergonzosas. Estos ejemplos hablan de una forma distintiva de participación política en Brasil, donde el humor y la crítica son expresiones populares de compromiso político entre los ciudadanos.

Figura 2.2. **Publicaciones de la precandidata a la alcaldía de São Paulo, Tabata Amaral, en Instagram**



Fuente: Resk, F., «Tabata usa deep fake para expor Nunes como Ken de Barbie e refaz post», Metropoles, 5 de abril de 2024, <https://www.metropoles.com/sao-paulo/tabata-deep-fake-nunes-ken-barbie>, fecha de consulta: 17 de junio de 2025.

Estos casos provocaron un debate sobre los desafíos de la aplicación uniforme de las políticas. Aunque la Resolución 23.732, emitida por el Tribunal Superior Electoral de Brasil en 2024, establece reglas claras para el uso de la IA en la propaganda electoral (TSE, 2024), un estudio del Laboratorio de Gobernanza y Regulación de la Inteligencia Artificial (Laboratório de Governança e Regulação de Inteligência Artificial), vinculado al Instituto Brasileño de Educación, Desarrollo e Investigación (Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa), concluyó que las sentencias de los tribunales locales de todo el país carecían de coherencia en la clasificación de los *deepfakes* a pesar de la claridad de la resolución. Por ejemplo, la denuncia del Movimiento Democrático Brasileño en relación con el video de Ken (mencionado anteriormente) fue finalmente desestimada por el tribunal electoral de São Paulo, que dictaminó que carecía de intención ilícita y entraba en el ámbito de la libertad de expresión. Sin embargo, otros tribunales de todo el país adoptaron un enfoque más estricto, prohibiendo cualquier uso de *deepfakes,* incluso cuando estaban etiquetados o eran claramente identificables (Caballero et al., 2024).

2.2. LA IA COMO AMPLIFICADOR DE LA VIOLENCIA POLÍTICA DE GÉNERO

Aunque México eligió por primera vez a una mujer presidenta, lo cual representa un avance en la representación política de las mujeres, en las elecciones de 2024 también se registró un aumento significativo de la violencia política de género. Datos del Instituto Nacional Electoral indican que durante el período electoral se registraron 205 denuncias relacionadas con este tipo de violencia, de las cuales el 42 % correspondieron a ataques en redes sociales. La violencia incrementó su frecuencia y gravedad a medida que se acercaba el día de las elecciones, pasando de los formularios digitales a las amenazas físicas, las agresiones e incluso los asesinatos (Rangel, 2024).

Uno de los casos más destacados fue el de la senadora del estado de Chihuahua Andrea Chávez, que denunció en las redes sociales la distribución de una imagen manipulada por IA que mostraba su rostro en el cuerpo semidesnudo de otra persona. Un informe de la Fundación Friedrich Naumann y del Consejo Alemán de Relaciones Exteriores (Del Pozo y Arroyo, 2024) también menciona casos en los que se utilizó la IA para "embellecer" a las candidatas. Si bien dicha manipulación de imágenes no constituye necesariamente violencia política de género, refuerza los estereotipos de la cosificación femenina en posiciones de poder y también representa un desafío para la aplicación de las leyes nacionales de México con respecto a la tecnología y la violencia digital de género. Como se ha mencionado anteriormente en este informe, si bien la llamada Ley Olimpia se diseñó principalmente para abordar la distribución no consentida de imágenes íntimas, los debates recientes sobre la ley han tomado en consideración su posible aplicación a contenidos generados por IA, como los desnudos *deepfake* (Borchardt, 2024)

Más allá de la exposición, el ridículo y la vergüenza asociada a este tipo de ataques, estos casos también causan daños materiales a los afectados, que se ven obligados a dedicar tiempo y esfuerzo a denunciar los incidentes, encausar a los responsables o hacer que se elimine el contenido. Durante las campañas electorales, el tiempo y el esfuerzo dedicados a combatir los *deepfakes* también pueden perjudicar las posibilidades electorales de un candidato. Si bien los políticos varones también son víctimas de ataques y burlas generados por la IA, se detecta un claro patrón de género cuando se examinan los casos que involucran a mujeres, generalmente con el objetivo de avergonzarlas y silenciarlas mediante la activación de estereotipos, relacionados sobre todo con la sexualidad.

Las elecciones de Brasil y México también demostraron la posibilidad de aplicar esta tecnología al fortalecimiento de la democracia y la participación ciudadana.

2.3. HERRAMIENTAS PARA LA PROMOCIÓN DE LA CIUDADANÍA

A pesar de los desafíos que plantea la inteligencia artificial en contextos electorales, las elecciones de Brasil y México también demostraron la posibilidad de aplicar esta tecnología al fortalecimiento de la democracia y

la participación ciudadana. Un claro ejemplo de ello fue el uso de *chatbots* y otros recursos generados por IA para fomentar una interacción más estrecha entre candidatos y votantes.

Aunque no es algo exclusivo de las elecciones de 2024 en América Latina, las elecciones municipales de Brasil proporcionaron un ejemplo positivo de cómo se pueden aprovechar los *chatbots* para mejorar la comunicación política y la participación de los votantes. Su despliegue fue particularmente notable entre los candidatos a la alcaldía de las principales capitales brasileñas, que utilizaron herramientas impulsadas por IA para interactuar con los votantes, presentar propuestas y movilizar partidarios para eventos de campaña. El uso de *chatbots* en las elecciones de 2024 en Brasil estaba regulado por el Tribunal Superior Electoral, que ordenó que los candidatos no pudieran simular personas reales y se indicara claramente que las interacciones estaban impulsadas por IA (Tribunal Superior Eleitoral, 2024).

Incluso dentro de los límites impuestos por las normas electorales, estas herramientas tienen un potencial significativo para agregar valor a la imagen de un candidato. Al automatizar las interacciones y proporcionar información a los votantes, estos asistentes virtuales no solo ampliaron el alcance de la campaña, sino que también activaron elementos culturales y simbólicos, influyendo en la percepción pública de los candidatos.

En México, casos similares demostraron el potencial positivo de la IA. Durante la precampaña, Xóchitl Gálvez utilizó la IA generativa para crear una versión digital de sí misma, llamada iXóchitl y comunicarse con los votantes en videos de campaña en las redes sociales. Gálvez también afirmó que utilizó herramientas de IA para prepararse para los debates electorales, simulando preguntas de su oponente (Ormerod, 2023).

El Instituto Nacional de Elecciones (INE) también desplegó herramientas de IA durante las elecciones mexicanas para ayudar a combatir la desinformación, lanzando el *chatbot* Inés como una herramienta para que los usuarios denunciaran contenidos potencialmente engañosos y que estos fueran verificados por empresas especializadas (INE, 2024b).

Las agencias gubernamentales responsables de supervisar el proceso electoral de Brasil también adoptaron *chatbots* impulsados por IA. En el estado de Bahía, el Tribunal Electoral utilizó el *chatbot* Maia, que brindó apoyo a través de WhatsApp y Telegram para la presentación de informes financieros de campañas y facilitó el acceso de los ciudadanos a la información relacionada con las elecciones, promoviendo una mayor transparencia y accesibilidad. En Pernambuco, otro estado brasileño, el asistente virtual Júlia permitió a los votantes acceder a información relacionada con las elecciones a través de Telegram, incluida la verificación de su estado de registro de votante, la obtención de su número de identificación de votante, la búsqueda de su lugar de votación y el acceso a orientación sobre la seguridad electoral y las regulaciones de campaña (Tribunal Regional Eleitoral–BA, 2024).

El uso de la IA en las elecciones de 2024 en Brasil y México se caracterizó principalmente por la experimentación relacionada con la creación de contenidos, y probablemente tuvo poco impacto en los resultados finales.

El uso de la IA en las elecciones de 2024 en Brasil y México se caracterizó principalmente por la experimentación relacionada con la creación de contenidos, y probablemente tuvo poco impacto en los resultados finales. En relación con esto, cabe destacar que, en varios casos, el engaño no fue la razón principal para el uso de IA. Entre los muchos riesgos imaginarios asociados a la IA, destaca como el más alarmante la posibilidad de generar contenidos imposibles de verificar. De cualquier manera, el impacto de la IA en la política y la vida cotidiana puede manifestarse de formas que van más allá del engaño, aumentando el potencial para construir discursos que ridiculicen y desacrediten.

En los casos de engaño, destacan los contenidos de baja tecnología (como las imágenes manipuladas) y el predominio de los formatos de audio, siendo estos últimos los que tienen mayor potencial para producir contenidos casi imposibles de verificar.

En cuanto a la producción de contenidos, también es importante tener en cuenta cómo se conecta el uso de la IA con dinámicas preexistentes, como se ve en los vídeos en los que los candidatos políticos respaldan plataformas de estafas financieras. En otras palabras, la IA mejora el impacto y el alcance de una estafa existente, aprovechando la visibilidad y el atractivo de los candidatos durante el período electoral para atraer a nuevas víctimas.

Los casos latinoamericanos analizados revelan que, si bien varios candidatos fueron objetivo de contenidos audiovisuales generados por IA, las candidatas fueron objeto específico de ataques relacionados con la violencia de género. En este contexto, el análisis indica que los grupos más vulnerables tienden a experimentar los impactos negativos del uso indebido de la IA con mayor intensidad, exacerbando las líneas preexistentes de discriminación y prejuicio. La ausencia de regulaciones específicas que aborden la producción de *deepfakes* para el acoso a mujeres y personas LGBTQIA+ en las propuestas legislativas de los países latinoamericanos analizados agrava aún más este problema, dada la recurrencia de este tipo de prácticas.

Si bien las aplicaciones de IA se han utilizado en gran medida para socavar la integridad del sistema de información electoral, también se han empleado como una herramienta para mejorar la ciudadanía —como en el caso de los *chatbots* creados por los tribunales electorales regionales en Brasil— y como un instrumento que profundiza la reflexión y el conocimiento político en el contexto electoral, ejemplificado por los experimentos de la candidata mexicana Xóchitl Gálvez con un avatar digital y el entrenamiento asistido por IA para debates televisados.

Por último, es importante destacar el carácter experimental e innovador de la popularización de las herramientas de IA en la vida cotidiana. Recientemente, en Brasil, el expresidente Bolsonaro aprovechó la confianza pública en las herramientas de IA generativa al pedir a los ciudadanos que usaran la IA para cuestionar la conducta de un juez de la Corte Suprema en un caso que acusaba a Bolsonaro de planear un golpe de Estado (Poder360, 2025). Este

acontecimiento tuvo lugar fuera del período electoral y queda fuera del alcance de este estudio, pero apunta a una forma de utilizar las herramientas de IA generativa no identificada previamente, como un mediador supuestamente neutral en conflictos y disputas sociales.

Capítulo 3

POLÍTICAS Y ELECCIONES DE IA EN AMÉRICA LATINA

Este capítulo ofrece una visión general del panorama actual de las políticas de IA, tanto a nivel mundial como en América Latina, y destaca los debates en torno al el papel de esta tecnología durante las elecciones y su impacto en la perspectiva de género y la inclusión en los marcos y estrategias de políticas. El capítulo se divide en dos subsecciones: la primera abarca los documentos publicados por organizaciones internacionales y la segunda por los gobiernos y las coaliciones regionales de América Latina.

3.1. **MARCO ANALÍTICO**

El marco analítico se elaboró para entender en qué medida los actores estatales y las organizaciones internacionales están abordando actualmente cuestiones relacionadas con la IA, tanto en el ámbito electoral como en el de género. Dado que se trata de problemas complejos y multifacéticos, el marco tiene por objeto establecer una escala que permita analizar si se están abordando diferentes facetas del fenómeno. Lo hace dividiendo los dos áreas—la IA y las elecciones y la IA y el género— en subconjuntos de aspectos relevantes para cada caso. Por ejemplo, la IA y el género se examinan analizando cómo los actores abordan la violencia de género facilitada por la tecnología a través de los *deepfakes*, la brecha de la IA, los efectos negativos de las tecnologías digitales en la participación política y los sesgos inducidos por los conjuntos de datos. Las facetas elegidas se ponen en práctica en las correspondientes preguntas cerradas, dirigidas a los marcos y estrategias de políticas internacionales y nacionales, como la siguiente: "¿Reconoce el texto la dimensión de género de la IA o la brecha digital?". Hay tres respuestas posibles: "Sí", "En parte" o "No". El nivel en que se cumple cada criterio indica a su vez si se está abordando actualmente la cuestión más amplia, y en qué aspectos.

La primera matriz (Figura 3.1) evalúa tres aspectos del uso de la IA en la distribución de contenidos en línea durante las elecciones: a) el uso de la IA para difundir desinformación política y contaminar el entorno de la información, poniendo en peligro la integridad de la información electoral; b) el diseño de sistemas de IA y los incentivos que repercuten en el entorno de la información; y c) instrumentos para mitigar el efecto perturbador de la IA en las elecciones.

Figura 3.1. **Matriz analítica sobre IA en contextos electorales, aplicada a documentos de política global y regional**

Organización Documento	**¿Contiene el texto información sobre la desinformación electoral y la integridad de las elecciones?**		**¿Contiene el texto información sobre modelos de negocio de IA con ánimo de lucro, mecanismos de supervisión humana y sistemas de recomendación?**			**¿Contiene el texto información sobre los sistemas de moderación de contenidos de la IA y su procedencia?**	
	Desinformación Electoral	**Integridad Electoral**	**Modelos de Negocio**	**Mecanismos de Supervisión Humana**	**Algoritmos de Recomendación y Clasificación**	**Sistemas de Moderación**	**Procedencia**

Fuente: Elaboración propia.

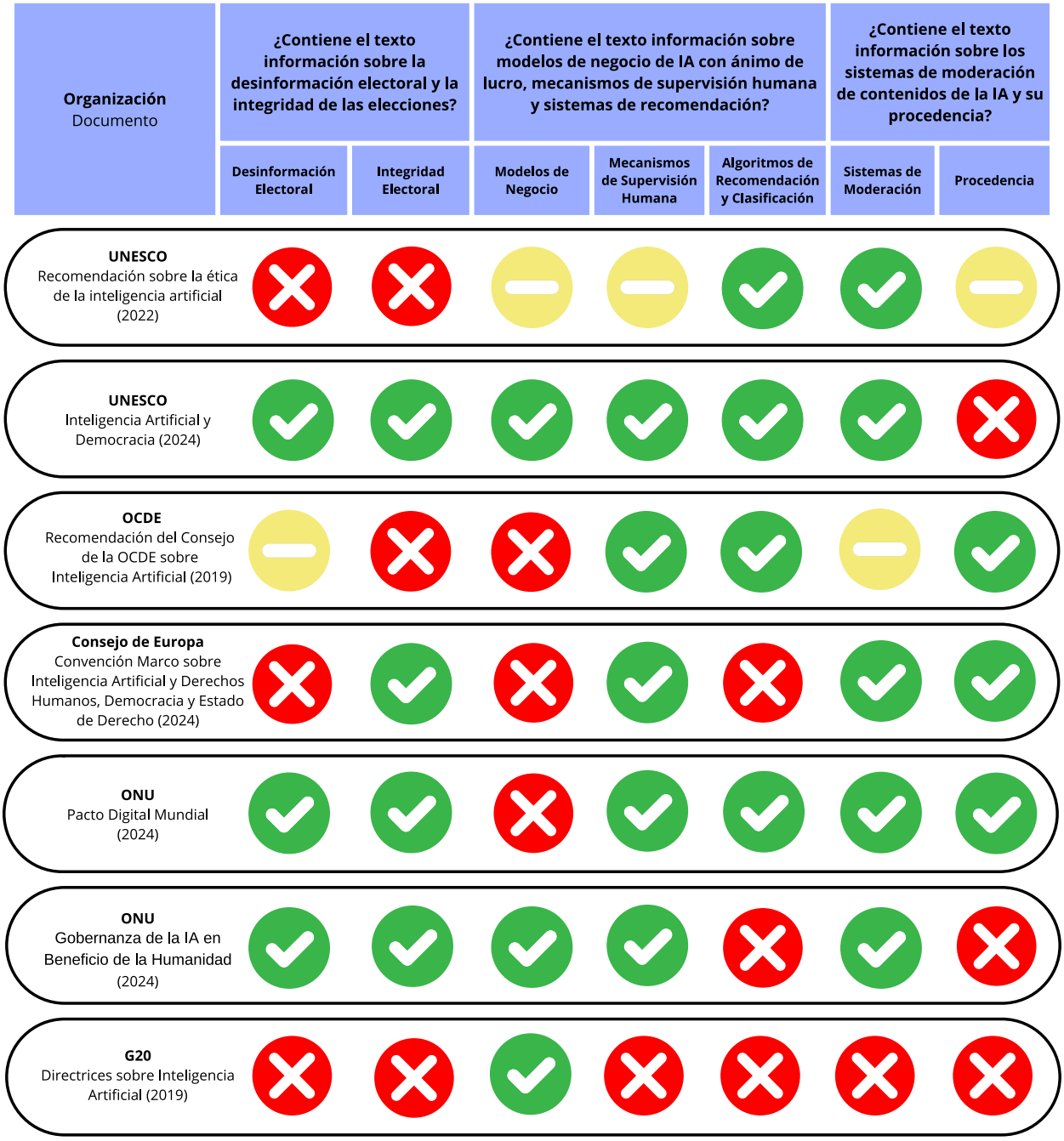
La segunda matriz (figura 3.2) se ocupa del impacto de la IA en las mujeres, las niñas y las comunidades LGBTQIA+, identificando cuatro facetas de la marginación digital: (a) el impacto significativo y desproporcionado de la IA generativa sobre las mujeres, las niñas y las personas LGBTQIA+; (b) el impacto negativo de las tecnologías de IA en la participación política de las mujeres, las niñas y las personas LGBTQIA+; (c) la transposición de la brecha digital de género a una brecha de IA, en la que las mujeres, las niñas y las personas LGBTQIA+ se enfrentan a disparidades en lo que respecta al conocimiento, la competencia y la confianza en los sistemas de IA; y (d) si las estrategias de IA abordan o no los sesgos discriminatorios contra las mujeres, las niñas y las personas LGBTQIA+.

Figura 3.2. **Matriz analítica sobre IA, género e inclusión, aplicada a documentos de políticas globales y regionales**

Organización Documento	**¿Aborda el texto el uso de deepfakes para acosar a mujeres y personas LGBTQIA+?**	**¿Se plantea en el texto el posible impacto negativo de la IA en la participación política de las mujeres y las personas LGBTQIA+?**	**¿Reconoce el texto la dimensión de género de la IA o la brecha digital?**	**¿Aborda el texto la cuestión de la discriminación a mujeres y las personas LGBTQIA+ inducida por las bases de datos usadas por los sistema de Inteligencia Artificial?**

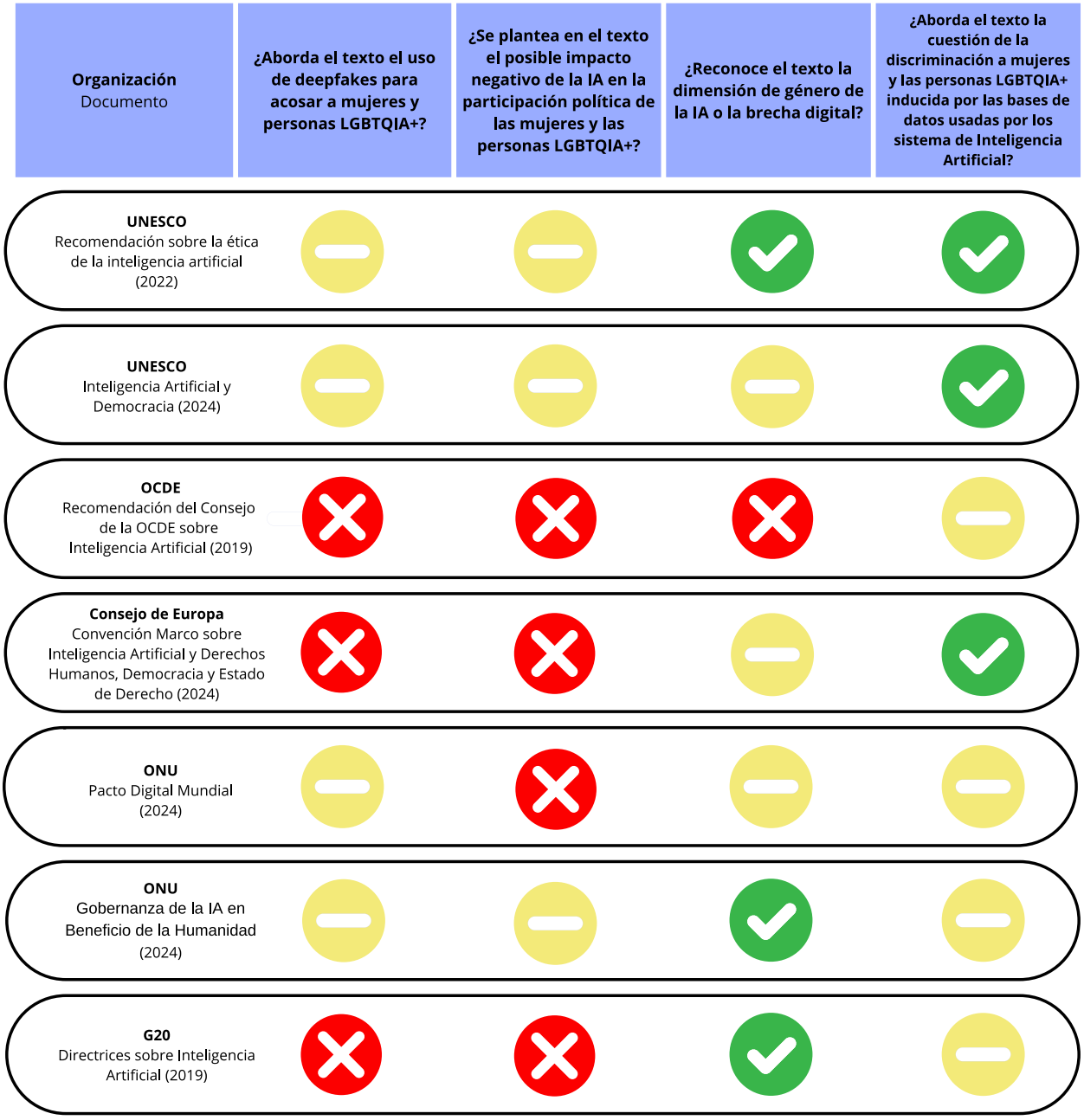
Fuente: Elaboración propia.

Figura 3.3. **Resumen del debate sobre la IA en contextos electorales en documentos de política mundial**



Fuente: United Nations Educational, Scientific and Cultural Organization (UNESCO), 'Recommendations on the Ethics of Artificial Intelligence', 16 de mayo de 2023; UNESCO, 'Artificial Intelligence and Democracy', 16 de mayo de 2023; Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence, adoptada el 22 de mayo de 2019, modificada el 3 de mayo de 2024, OECD Legal Instruments; Council of Europe, 'Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law', 5 de septiembre de 2024; United Nations, Global Digital Compact, 22 de septiembre de 2024, A/79/L.2; United Nations, *Governing AI for Humanity: Final Report* (New York: United Nations, 2024); and G20, Ministerial Statement on Trade and Digital Economy (AI Principles), 9 de junio de 2019.

Figura 3.4. **Resumen del debate sobre la IA, el género y la inclusión en los documentos de políticas mundiales**



Fuente: United Nations Educational, Scientific and Cultural Organization (UNESCO), 'Recommendations on the Ethics of Artificial Intelligence', 16 de mayo de 2023; UNESCO, 'Artificial Intelligence and Democracy', 16 de mayo de 2023; Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence, adoptada el 22 de mayo de 2019, modificada el 3 de mayo de 2024, OECD Legal Instruments; Council of Europe, 'Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law', 5 de septiembre de 2024; United Nations, Global Digital Compact, 22 de septiembre de 2024, A/79/L.2; United Nations, *Governing AI for Humanity: Final Report* (New York: United Nations, 2024); and G20, Ministerial Statement on Trade and Digital Economy (AI Principles), 9 de junio de 2019.

3.2. **ANÁLISIS GLOBAL**

En todos los marcos analizados se evalúan al menos algunos de los indicadores operacionales seleccionados para la IA en contextos electorales, con una tasa de cobertura explícita acumulada del 55,1 %. Sin embargo, no hay una pregunta concreta que se aborde con más frecuencia o se omita por completo. El punto planteado con mayor frecuencia fue la necesidad de incluir mecanismos de supervisión humana en el desarrollo y despliegue de la IA, abordado explícitamente por el 71,4 % de los marcos y parcialmente por el 14,3 %. En cambio, tema menos tratado en los documentos es el de los modelos de negocio con fines de lucro de las empresas de IA y sus incentivos asociados, de los que se ocupa explícitamente el 42,8 % de los textos y parcialmente el 14,3 %. En particular, el documento que contiene la menor cantidad de discusión de los indicadores por un amplio márgen es “Directrices de Inteligencia Artificial” del G20, que se centra únicamente en los modelos de negocio con fines de lucro de las empresas de IA.

La falta de cohesión entre los marcos amenaza con socavar el impacto multifacético de la IA, subestimando la complejidad y diversidad de sus riesgos.

Si bien todas las políticas, recomendaciones y directrices examinadas abordan en cierta medida la interacción de la IA y el entorno de la información electoral, las facetas y la complejidad de los debates varían enormemente. Esta variación puede atribuirse, en parte, al alcance de los documentos, ya que los textos más completos dedican espacio a problematizar cuestiones en la intersección de la IA y su impacto en la sociedad. Esta discrepancia puede ejemplificarse en los extensos debates sobre la integridad de la información electoral que se encuentran en el informe de la UNESCO sobre *Inteligencia Artificial y Democracia* (2024) y en el informe de la ONU sobre la *Gobernanza de la IA en Beneficio de la Humanidad* (2024), en comparación con los limitados ejemplos proporcionados en las Directrices de IA del G20 (2019). Además, los instrumentos con un propósito integral, como el Pacto Digital Mundial (*Global Digital Compact*) de las Naciones Unidas, carecen de debate sobre los mecanismos e incentivos subyacentes decisivos para lograr una visión completa de por qué y cómo la IA da lugar a ciertas amenazas para el entorno informativo. La falta de cohesión entre los marcos pone en riesgo el abordaje integral del impacto multifacético de la IA, subestimando la complejidad y diversidad de sus riesgos.

En general, los textos incluyen una cobertura significativamente menos amplia de las cuestiones de género, con una tasa de cobertura explícita acumulada del 21,4 % (una tasa de parcial del 50 %). El problema que se aborda con más frecuencia es la existencia de sesgos contra las mujeres y las personas LGBTQIA+ inducidos por conjuntos de datos, que se aborda explícitamente en el 42,9 % de los textos y parcialmente en el 57,1 %. El posible impacto negativo de la IA en la participación política de las mujeres y las personas LGBTQIA+ es el aspecto menos abordado por el marco analítico, ya que solo el 42,9 % de los marcos plantean la cuestión. Casi ninguno de los textos abarcaba explícitamente más de uno de los indicadores, y la Recomendación del Consejo de Inteligencia Artificial de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) era el que menos, con un solo punto cumpliéndose parcialmente.

En cuanto a las cuestiones de género, los textos muestran niveles similares de incongruencia en cuanto a los temas que se abordan y quiénes se incluyen en los grupos afectados. En general, las mujeres y las niñas se identifican con frecuencia entre los grupos que se ven afectados de manera desproporcionada por la IA. Es mucho menos común que los miembros de la comunidad LGBTQIA+ sean identificados explícitamente como particularmente vulnerables. La representación incompleta de los grupos vulnerables a la violencia de género facilitada por la tecnología corre el riesgo de pasar por alto el daño y las necesidades específicas de las personas LGBTQIA+, perpetuando así las estructuras de opresión existentes en los espacios en línea.

3.3. **ANÁLISIS REGIONAL: AMÉRICA LATINA**

En general, los documentos de política pública latinoamericana analizados no abordan la mayoría de los indicadores, con una tasa acumulada de respuesta afirmativa del 26,5 %, 28,6 puntos porcentuales menos que la de los documentos globales. Dos documentos (la Estrategia Nacional de Inteligencia Artificial del Perú y la Declaración de Montevideo sobre la Inteligencia Artificial y su Impacto en América Latina) no abordan un solo indicador. Por un amplio margen, el indicador más discutido es la necesidad de mecanismos de supervisión humana, que se plantea en el 71,4 % de los textos analizados. Los puntos menos discutidos son los sistemas de procedencia, que no se mencionan en ninguno de los textos, seguidos de la desinformación electoral, la integridad electoral y los modelos de negocio de IA, cada uno cubierto en un solo documento.

En comparación con las políticas globales analizadas, las estrategias regionales en América Latina abordan con poca frecuencia y de manera inadecuada la conexión entre la IA y la distribución y gestión de contenidos en línea. Aparte de las escasas menciones a la necesidad de emplear la supervisión humana y la moderación de contenido en línea, los documentos no abordan la posible influencia de la IA en los entornos de información electoral, ni abogan por mecanismos específicos para administrar la IA, como los sistemas de procedencia o la transparencia en los algoritmos de recomendación y jerarquización.

Varios de los países investigados han negociado y respaldado oficialmente acuerdos globales de política digital, como la participación activa de Perú y México en el desarrollo del Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho. Dado que el desarrollo de políticas de IA es un proceso mundial en curso, las incoherencias o las lagunas entre las estrategias nacionales, regionales y globales pueden explicarse en parte por la superposición de muchos procesos que tienen lugar en paralelo y que supuestamente se informan entre sí. Sin embargo, esta explicación no es totalmente representativa, ya que las estrategias nacionales publicadas más recientemente, como la de Costa

Figura 3.5. **Panorama de la discusión sobre la IA en contextos electorales en documentos de política latinoamericana**

Fuente: Argentina, Subsecretaría de tecnologías de la información, Recomendaciones para una inteligencia artificial fiable [Recommendations for reliable artificial intelligence], 1 de junio de 2023; Brazil, Ministério da Ciência, Tecnologia e Inovações, Estratégia Brasileira de Inteligência Artificial [Artificial Intelligence Strategy of Brazil], 13 de julio de 2021; Chile, Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, Política nacional de inteligencia artificial [National Artificial Intelligence Policy], 27 de octubre de 2021; Costa Rica, Ministry of Science, Innovation, Technology, and Telecommunications (MICITT), *National Artificial Intelligence Strategy of Costa Rica 2024–2027* (San José: MICITT, 2024); Peru, Secretaría de Gobierno y Transformación Digital, Estrategia Nacional de Inteligencia Artificial [National Artificial Intelligence Strategy], 2021; IA2030Mx Coalition, Agenda Nacional Mexicana de Ineligencia Artificial [Mexican National Agenda on Artificial Intelligence], septiembre de 2020; y Montevideo Declaration on Artificial Intelligence and its Impact in Latin America, 10 de marzo de 2023.

Rica (2024), aún no reconocen cuestiones clave de la IA y la integridad de la información.

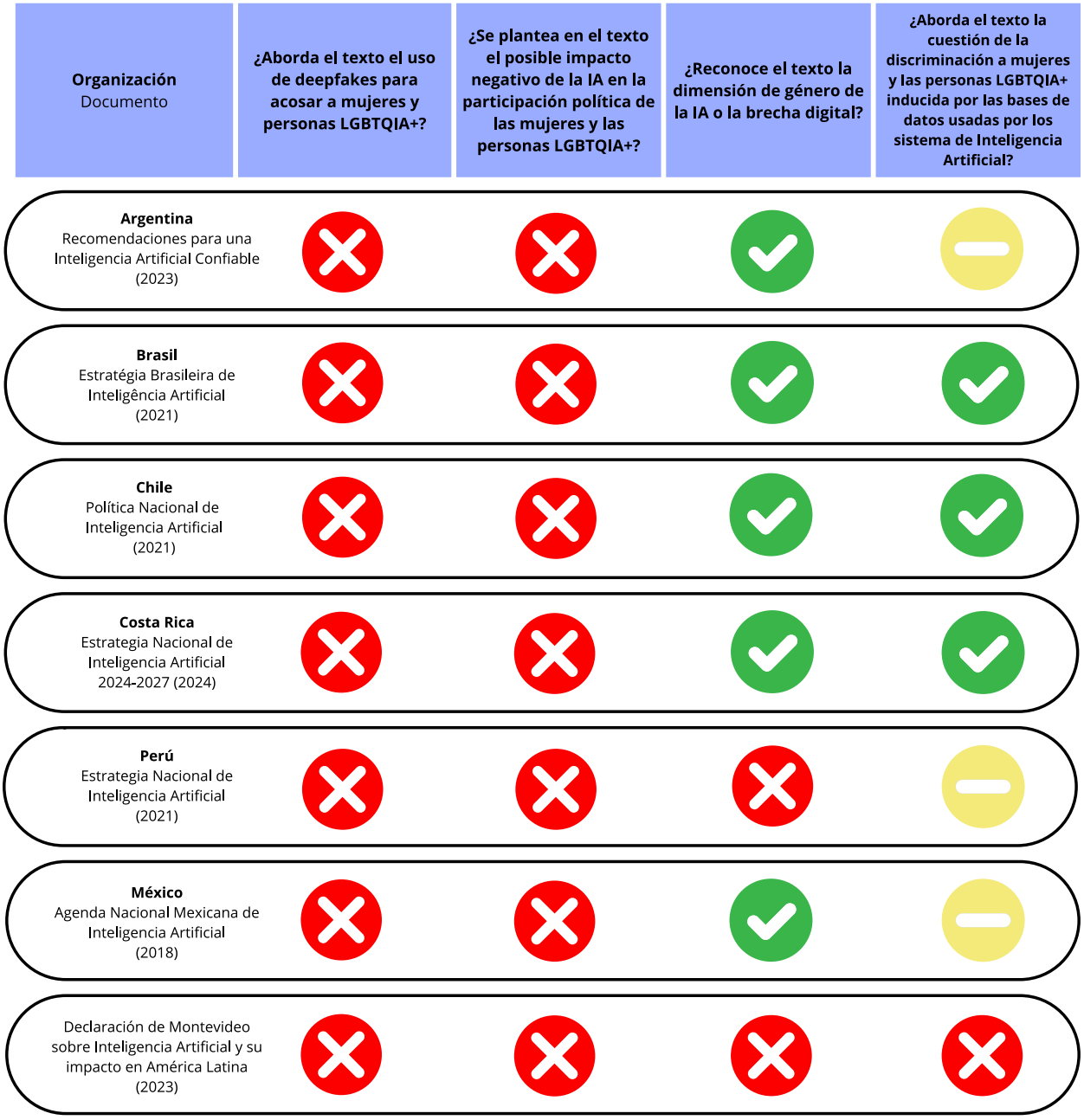
Si bien los documentos analizados también carecían de discusiones sobre temas relacionados con la IA y el género, hay patrones más claros en las áreas que se abordan. La tasa total de cobertura explícita se sitúa en el 28,6 %; todos los puntos tratados se refieren a la IA de género o a la brecha digital y a la existencia de sesgos estructurales y discriminatorios de género. Ninguno de los documentos analiza cómo se utilizan los *deepfakes* como una forma de violencia de género contra las mujeres facilitada tecnológicamente o el posible impacto negativo de la IA en la participación política de las mujeres y las personas marginadas de género. Por el contrario, la mayoría de los textos abordan explícita o parcialmente la dimensión de género de la IA o la brecha digital (71,4 % de menciones explícitas) y la cuestión de los sesgos de género inducidos por los conjuntos de datos (71,4 % de menciones explícitas, 42,8 % de menciones parciales).

Al igual que sus equivalentes globales, las estrategias de IA de América Latina presentan un mosaico de declaraciones sobre cómo la IA impacta en los grupos marginados por motivos de género. Si bien a menudo se reconoce el sesgo discriminatorio sistemático, la forma en la que el sesgo de género afecta a las personas LGBTQIA+ no se aborda de manera consistente, como en los casos de Argentina, Costa Rica y México. Sobre todo, los textos carecen de debates sobre cómo la IA generativa utilizada como arma se dirige de manera desproporcionada contra las mujeres, las niñas y las personas LGBTQIA+ y cómo podría interrumpir o desalentarlas de la participación política. Casi todas las estrategias analizadas abordan cómo la brecha digital limita el acceso de los grupos marginados, en particular de las mujeres, a las tecnologías de IA. En la mayoría de los casos, las declaraciones sobre la brecha digital se traducen en recomendaciones de medidas para aumentar la participación de las mujeres en el desarrollo de la IA y, en general, en los campos CTIM (Ciencia, Tecnología, Ingeniería y Matemáticas). Sin embargo, el impacto potencial que la división de la IA podría tener en la participación cívica está ausente, dejando de lado cómo las comunidades marginadas no solo carecen de acceso a los beneficios potenciales de la IA, sino que también son objetivos activos cuando se utiliza para hacer daño.

Las estrategias de IA de América Latina presentan un mosaico de declaraciones sobre cómo la IA impacta en los grupos marginados por motivos de género. Si bien a menudo se reconoce el sesgo discriminatorio sistemático, la forma en que el sesgo de género afecta a las personas LGBTQIA+ no se aborda de manera consistente.

En el caso brasileño, avances políticos recientes han ampliado el enfoque regulatorio y estratégico del país hacia la IA. En 2024, relanzó su estrategia nacional de IA, incorporando inversiones públicas adicionales y haciendo hincapié en iniciativas como el desarrollo de un modelo extenso de lengua portuguesa y la aplicación de la IA en los servicios públicos, con inversiones previstas hasta 2028 (MCTI, 2024). Además, en diciembre de 2024, el

Figura 3.6. **Panorama de la discusión sobre IA, género e inclusión en los documentos de políticas de América Latina**



Fuente: Argentina, Subsecretaría de tecnologías de la información, Recomendaciones para una inteligencia artificial fiable [Recommendations for reliable artificial intelligence], 1 de junio de 2023; Brazil, Ministério da Ciência, Tecnologia e Inovações, Estratégia Brasileira de Inteligência Artificial [Artificial Intelligence Strategy of Brazil], 13 de julio de 2021; Chile, Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, Política nacional de inteligencia artificial [National Artificial Intelligence Policy], 27 de octubre de 2021; Costa Rica, Ministry of Science, Innovation, Technology, and Telecommunications (MICITT), *National Artificial Intelligence Strategy of Costa Rica 2024–2027* (San José: MICITT, 2024); Peru, Secretaría de Gobierno y Transformación Digital, Estrategia Nacional de Inteligencia Artificial [National Artificial Intelligence Strategy], 2021; IA2030Mx Coalition, Agenda Nacional Mexicana de Ineligencia Artificial [Mexican National Agenda on Artificial Intelligence], septiembre de 2020; y Montevideo Declaration on Artificial Intelligence and its Impact in Latin America, 10 de marzo de 2023.

Senado brasileño aprobó el Marco Legal de la Inteligencia Artificial (*Marco Legal da Inteligência Artificial*), que establece estructuras de gobernanza para el desarrollo y la comercialización de la IA (Zanatta y Rielli, 2024; Alerta de Política Digital, s.f.). El marco también introduce disposiciones para salvaguardar los derechos de autor de los creadores cuyas obras se utilizan en el entrenamiento de la IA. Sin embargo, la legislación elude en gran medida cuestiones relacionadas con las plataformas de redes sociales, la moderación de contenidos o las elecciones, temas que se consideraron como muy polémicos entre los actores políticos, en particular los partidos de derecha.

Capítulo 4

RECOMENDACIONES DE POLÍTICA

1. *Abordar los usos de alto riesgo de la IA en los procesos electorales.* Los marcos normativos mundiales y regionales de IA deben abordar de forma proactiva las aplicaciones controvertidas y dañinas en lugar de soslayar los debates sobre usos malintencionados o políticamente perturbadores. Muchos de los desafíos planteados por la IA en las elecciones de México no estaban adecuadamente cubiertos por las regulaciones nacionales, regionales o globales, en particular en lo que respecta a la manipulación electoral y la violencia digital de género. En cambio, a pesar de los desafíos en la aplicación cohesiva, la regulación judicial preventiva de Brasil por parte del Tribunal Superior Electoral ofrece vías concretas para gestionar las aplicaciones directas de IA, como prohibir el contenido generado por IA sin etiquetar y permitir interacciones automatizadas reguladas impulsadas por IA.

2. *Exigir la transparencia en los sistemas de IA utilizados para la curación de contenidos y la publicidad política.* Los gobiernos deben exigir a las plataformas que divulguen los conjuntos de datos que utilizan para entrenar los algoritmos de IA, expliquen la lógica detrás de la jerarquización y personalización del contenido y proporcionen explicaciones claras de por qué a los usuarios se les muestran anuncios o contenido político específicos.

3. *Desarrollar mecanismos de procedencia y etiquetado para los contenidos generados por IA.* Los países y las plataformas deben implementar sistemas sólidos de procedencia y exigir el etiquetado del material generado por IA, especialmente durante los períodos electorales, para garantizar que los usuarios puedan verificar el origen y la autenticidad del contenido en línea.

4. *Ampliar el enfoque de las políticas más allá de la manipulación política.* El papel de la IA en la política va más allá de las tácticas tradicionales de desinformación, influyendo en la movilización de campañas, la

participación basada en el humor y la explotación económica. La IA se utilizó durante las elecciones de 2024 en Brasil y México de muchas maneras que no eran explícitamente engañosas; más bien, ayudaban a dar forma a los relatos políticos a través del humor, las representaciones gráficas o las estrategias de contenido monetizado. Las políticas futuras deben reconocer la evolución del papel de la IA en la dinámica electoral, garantizando la rendición de cuentas no solo por la manipulación deliberada, sino también por sus efectos más amplios en la participación política y los ecosistemas digitales.

5. *Abordar la violencia de género facilitada por la IA en los marcos jurídicos y normativos.* Las estrategias nacionales de IA deben reconocer y abordar explícitamente los *deepfakes* y otras formas de abuso facilitadas por la tecnología que afectan de manera desproporcionada a las mujeres, las niñas y las personas LGBTQIA+, especialmente en contextos políticos y electorales.

6. *Introducir auditorías algorítmicas y garantizar el acceso a los datos para una supervisión independiente.* Las políticas públicas deben garantizar que los organismos independientes, incluidas las comisiones electorales y las organizaciones de la sociedad civil, tengan acceso a los datos y algoritmos de las plataformas con fines de auditoría, especialmente durante las elecciones, para evaluar el sesgo, la discriminación y los riesgos democráticos.

7. *Garantizar que las estrategias de IA promuevan tanto la protección como la participación de los grupos marginados.* Los gobiernos deben ir más allá del reconocimiento de la brecha digital y promover activamente la inclusión de las comunidades indígenas, las mujeres y las personas LGBTQIA+ en el diseño, el desarrollo y la gobernanza de la IA, incluidos los mecanismos para la soberanía indígena de los datos y la equidad lingüística.

8. *Alentar a los partidos políticos, las campañas políticas y los candidatos a firmar un código de conducta.* Aunque los códigos de conducta son una forma de derecho sin sanción [*soft law*], pueden fomentar la cooperación y campañas digitales positivas que eviten los usos negativos de las herramientas digitales para manipular los ecosistemas de información. A través de códigos de conducta, los partidos políticos, las campañas y los candidatos pueden acordar la limitación del posible impacto de las tecnologías digitales, especialmente la IA, adoptar buenas prácticas y proporcionar espacios más seguros para *campañas políticas saludables*.

9. *Detener la monetización de la manipulación de la información y la falta de controles de financiamiento político para las campañas digitales.* La manipulación del entorno informativo suele ser rentable y puede serlo aún más con la ayuda de la IA. Los gobiernos, las plataformas digitales y los partidos políticos deben abordar la monetización de la manipulación

de la información para reducir los incentivos financieros para participar en prácticas manipuladoras. Del mismo modo, la falta de supervisión de las finanzas políticas para las campañas digitales a menudo permite que partidos y candidatos eludan la regulación financiera mientras hacen campaña en línea.

Referencias

Aaberg, R., Gurung, P., Kandie, C., Méndez, L. V., Xón, M. V., *Engaging Indigenous Peoples in Elections* [Involucrar a los pueblos indígenas en las elecciones] (Washington D.C.: International Foundation for Electoral Systems, 2024), <https://www.ifes.org/sites/default/files/2024-02/IFES%20Engaging%20Indigenous%20Peoples%20in%20Elections.pdf>, fecha de consulta: 18 de junio de 2025.

Ali, M., Mislove, A. Sapiezynski, P., Rieke, A., Korolova, A., "Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging" [Algoritmos de distribución de anuncios: los árbitros ocultos de los mensajes politicos], Actas de la 14ª Conferencia Internacional de la ACM sobre Búsqueda Web y Minería de Datos, 17 de diciembre de 2019, <https://www.semanticscholar.org/reader/822a9eed86d0375719ce71f0c2ef3b09feb458f5>, fecha de consulta: 18 de junio de 2025.

Amnesty International, "Surveillance giants: How the business model of Google and Facebook threatens human rights" [Gigantes de la vigilancia: cómo el modelo de negocio de Google y Facebook amenaza los derechos humanos], 21 de noviembre de 2019, <https://www.amnesty.org/en/documents/pol30/1404/2019/en>, fecha de consulta: 2 de junio de 2025.

ARTICLE 19, *Protecting Freedom of Expression Through the Unbundling of Services, Open Markets, Competition, and Users' Empowerment* [Protección de la libertad de expresión mediante la desagregación de servicios, la apertura de mercados, la competencia y el empoderamiento de los usuarios] (London: ARTICLE 19, 2021), <https://www.article19.org/wp-content/uploads/2021/12/Taming-big-tech_FINAL_8-Dec-1.pdf>, fecha de consulta: 17 de junio de 2025.

Associated Press, "A violent, polarized Mexico goes to the polls to choose between 2 women presidential candidates" [Un México violento y polarizado acude a las urnas para elegir entre dos candidatas presidenciales], 31 de mayo de 2024, <https://apnews.com/article/mexico-election-polarized-divided-heat-violence-4d5f620f0f8f9b7ef6efa8b3083561a8>, fecha de consulta: 18 de junio de 2025.

Borchardt, C., "¿Qué dice la Ley Olimpia sobre el uso de la IA?", El Imparcial, 10 de octubre de 2024, <https://www.elimparcial.com/mexico/2024/10/11/que-dice-la-ley-olimpia-sobre-el-uso-de-la-ia>, fecha de consulta: 18 de junio de 2025.

Burgess, M., "The biggest deepfake abuse site is growing in disturbing ways" [El sitio de abuso deepfake más grande está creciendo de manera inquietante], WIRED, 15 de diciembre de 2021, <https://www.wired.com/story/deepfake-nude-abuse>, fecha de consulta: 18 de junio de 2025.

Caballero, B. I., Ammon, B., Gomes, C., da Silva, C. C., Gomes, E., Roquete, F., Dias, F., Maia, G., Lopes, J., Costa, J., Mendes, L. S., Teixeira Ferreira, L. M., de Paula Silveira, M., de Oliveira, M., Vogel, S., Junquilho, T. A. y Marcílio, T., *Construindo consensos:* Deep fakes *nas eleições de 2024* [Construyendo consensos: *deepfakes* en las elecciones de 2024] (Brasília: Instituto Brasiliense de Direito Público, 2024), <https://www.idp.edu.br/arquivos/cedis/IDP%20-%20LIA%2C%20CEDIS%20e%20ETHICS4AI%20-%20Nota%20T%C3%A9cnica%20-%20Construindo%20Consensos%20-%20Deep%20Fakes%20nas%20Elei%C3%A7%C3%B5es%20de%202024.pdf>, fecha de consulta: 2 de junio de 2025.

Carter Center, The, *Analyzing Bolivia's 2020 General Elections: Final Report* [Análisis de las elecciones generales de Bolivia de 2020: Informe final] (Atlanta: Carter Center, 2021), <https://www.cartercenter.org/resources/pdfs/news/peace_publications/election_reports/bolivia-2020-final-report.pdf>, fecha de consulta: 2 de junio de 2025.

Cerdán Martínez, V. y Padilla Castillo, G., "Historia del 'fake' audiovisual: 'deepfake' y la mujer en un imaginario falsificado y perverso", *Historia y Comunicación Social*, 24/2 (2019), págs. 505-520, <https://doi.org/10.5209/hics.66293>.

Coppock, A., Hill, S. J. y Vavreck, L., "The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments" [Los pequeños efectos de la publicidad política son pequeños independientemente del contexto, el mensaje, el emisor o el receptor: evidencia de 59 experimentos aleatorios en tiempo real], *Science Advances*, 6/26 (2020), <https://doi.org/10.1126/sciadv.abc4046>.

Del Pozo, C. M. y Arroyo, D. R., *Generative IA and Its Influence on Mexico's 2024 Elections* [La IA generativa y su influencia en las elecciones de 2024 en México] (Potsdam-Babelsberg: Friedrich Naumann Foundation y German Council on Foreign Relations, 2024), <https://dgap.org/system/files/article_pdfs/Policy%20Paper%20AI%20on%20Mexico%27s%202024%20Electons%20PDF.pdf>, fecha de consulta: 2 de junio de 2025.

Digital Policy Alert, "Brazil: Passed Bill No. 2338 of 2023 regulating the use of Artificial Intelligence including data protection measures" [Brasil: Aprobó el Proyecto de Ley n.º 2338 de 2023 que regula el uso de la Inteligencia Artificial, incluyendo medidas de protección de datos], [s. f.], <https://digitalpolicyalert.org/event/25132-passed-bill-no-2338-of-2023-regulating-the-use-of-artificial-intelligence-including-data-protection-measures>, fecha de consulta: 17 de junio de 2025.

DiResta, R., *Invisible Rulers: The People Who Turn Lies into Reality* [Gobernantes invisibles: Las personas que convierten las mentiras en realidad] (Nueva York: Public Affairs, 2024).

El Economista, "AMLO alerta sobre estafa con IA utilizando su imagen para invertir en Pemex: ¿Cuál es la inversión mínima que piden los estafadores?", 26 de abril de 2024, <https://www.eleconomista.com.mx/politica/AMLO-alerta-sobre-estafa-con-IA-utilizando-su-imagen-para-invertir-en-Pemex-Cual-es-la-inversion-minima-que-piden-los-estafadores-20240426-0049.html>, fecha de consulta: 18 de junio de 2025.

Elswah, M., "Investigating content moderation systems in the Global South" [Investigando los sistemas de moderación de contenidos en el Sur Global], Center for Democracy & Technology, 30 de enero de 2024, <https://cdt.org/insights/investigating-content-moderation-systems-in-the-global-south>, fecha de consulta: 17 de junio de 2025.

European Partnership for Democracy, "Targeting and amplification in online political advertising" [Segmentación y amplificación en la publicidad política online], [s. f.], <https://epd.eu/content/uploads/2023/08/Targeting-and-amplification-in-online-political-advertising.pdf>, fecha de consulta: 17 de junio de 2025.

Farrugia, B., "Brazil election: AI deepfakes" [Elecciones en Brasil: deepfakes con IA], DFRLab, 26 de noviembre de 2024, <https://dfrlab.org/2024/11/26/brazil-election-ai-deepfakes>, fecha de consulta: 2 de junio de 2025.

FGV Comunicação Rio, *Artificial Intelligence in the World: Mapping Its Uses During the 2024 Elections* [Inteligencia artificial en el mundo: mapeando sus usos durante

las elecciones de 2024] (Rio de Janeiro: FGV Comunicação Rio, 2024), <https://midiademocracia.fgv.br/en/estudos/inteligencia-artificial-no-mundo-mapeamento-dos-usos-durante-eleicoes-de-2024>, fecha de consulta: 17 de junio de 2025.

FGV DAPP, *Nem Tão Simples: O Impacto Da Inteligência Artificial No Ecossistema Informacional* [El impacto de la inteligencia artificial en el ecosistema informativo] (Rio de Janeiro: Fundação Getulio Vargas, 2017), <https://repositorio.fgv.br/server/api/core/bitstreams/4eb748af-90ff-4f93-a3dd-6ff58382580b/content>, fecha de consulta: 2 de junio de 2025.

Flores Saviaga, C. y Savage, S., "Generative AI misinformation in the 2024 Mexican elections" [Desinformación generativa de IA en las elecciones mexicanas de 2024], Series on Generative Artificial Intelligence and Elections, Center for Media Engagement, de octubre de 2024, <https://mediaengagement.org/research/generative-artificial-intelligence-and-elections>, fecha de consulta: 16 de junio de 2025.

Fundación Multitudes, *Análisis comparativo de los marcos legales actuales en América Latina para combatir la violencia de género en línea y apoyar el liderazgo de las mujeres en la vida pública* (Fundación Multitudes, 2023), <https://www.boell.de/sites/default/files/2023-11/estudio-comparativo-fm-es.pdf>, fecha de consulta: 17 de junio de 2025.

Fundacja Panoptykon, "Fixing Recommender Systems: From identification of risk factors to meaningful transparency and mitigation" [Reparar los sistemas de recomendación: de la identificación de factores de riesgo a una la transparencia y mitigación significativa], 2023, pág.3, <https://en.panoptykon.org/fixing-recommender-systems-identification-risk-factors-meaningful-transparency-and-mitigation>, fecha de consulta: 18 de julio de 2025.

Funk, A., Shahbaz, A. y Vesteinsson, K., *Freedom on the Net 2023: The Repressive Power of Artificial Intelligence* [Libertad en la red 2023: El poder represivo de la inteligencia artificial] (Washington, DC: Freedom House, [s. f.]), <https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence>, fecha de consulta: 2 de junio de 2025.

G20, AI Principles [Principios de IA], 2019, <https://wp.oecd.ai/app/uploads/2021/06/G20-AI-Principles.pdf>, fecha de consulta: 19 de junio de 2025.

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M. y Sedova, K., "Generative language models and automated influence operations: Emerging threats and potential mitigations" [Modelos de lenguaje generativo y operaciones de influencia automatizadas: amenazas emergentes y posibles mitigaciones], Georgetown University's Center for Security and Emerging Technology, OpenAI y Stanford Internet Observatory, enero 2023, <https://doi.org/10.48550/arXiv.2301.04246>.

González-Véliz, C. y Cuzcano-Chavez, X., "Desafíos y dimensiones de la desinformación en ALAC deepfakes y la urgencia de proteger los derechos de las mujeres", *Revista Espacio I+D Innovación más Desarrollo*, 13 (2024), págs. 162-175, <https://doi.org/10.31644/IMASD.36.2024.a19>.

Gorton, W. A., "Manipulating citizens: How political campaigns' use of behavioral social science harms democracy" [Manipulando a los ciudadanos: Cómo el uso de las ciencias sociales del comportamiento en las campañas políticas perjudica la democracia], *New Political Science*, 38/1 (2016), págs. 61-80. <https://doi.org/10.1080/07393148.2015.1125119>.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. y Lazer, D., "Fake news on Twitter during the 2016 U.S. presidential election" [Noticias falsas en Twitter

durante las elecciones presidenciales estadounidenses de 2016], *Science*, 363/6425 (2019), págs. 374-378, <https://www.science.org/doi/10.1126/science.aau2706>.

Hofmann, V., Kalluri, P. R., Jurafsky, D. y King, S., "Dialect prejudice predicts AI decisions about people's character, employability, and criminality" [El prejuicio dialectal predice las decisiones de la IA sobre el carácter, la empleabilidad y la criminalidad de las personas], 1 de marzo de 2024, <https://doi.org/10.48550/arXiv.2403.00742>.

IDEA Internacional, *Global State of Democracy 2024: Strengthening the Legitimacy of Elections in a Time of Radical Uncertainty* [Estado global de la democracia 2024: Fortalecimiento de la legitimidad de las elecciones en tiempos de incertidumbre radical] (Stockholm: IDEA Internacional, 2024), <https://doi.org/10.31752/idea.2024.55>.

Instituto Nacional Electoral (INE), "Programa de resultados electorales preliminares 2024", 2024a, <https://prep2024.ine.mx/publicacion/nacional/presidencia/nacional/candidatura>, fecha de consulta: 2 de junio de 2025.

—, "Te presentamos a 'Inés', la asistente virtual del INE para combatir la desinformación sobre las elecciones 2024", 26 de abril 2024, <https://centralelectoral.ine.mx/2024/04/26/te-presentamos-a-ines-la-asistente-virtual-del-ine-para-combatir-la-desinfomacion-sobre-las-elecciones-2024>, fecha de consulta: 17 de junio de 2025.

Juneja, P., *La inteligencia artificial para la gestión electoral* (Estocolmo: IDEA Internacional, 2024), <https://doi.org/10.31752/idea.2024.89>.

Kahn, G., "Generative AI is already helping fact-checkers. But it's proving less useful in small languages and outside the West" [La IA generativa ya ayuda a los verificadores de datos. Pero resulta menos útil en idiomas pequeños y fuera de Occidente], Reuters Institute for the Study of Journalism, 29 de abril de 2024, <https://reutersinstitute.politics.ox.ac.uk/news/generative-ai-already-helping-fact-checkers-its-proving-less-useful-small-languages-and>, fecha de consulta: 2 de junio de 2025.

Martínez Holguín, P., "Mexico now has 13 female governors after 2024 elections" [México ya tiene 13 gobernadoras tras las elecciones de 2024], San Diego Red, 4 de junio de 2024, <https://www.sandiegored.com/en/news/258798/Mexico-now-has-13-female-governors-after-2024-elections>, fecha de consulta: 18 de junio de 2025.

Mexico News Daily, "Video circulating of Claudia Sheinbaum is seem a 'deepfake'" [El video de Claudia Sheinbaum, que circula parece un deepfake], 26 de enero de 2024, <https://mexiconewsdaily.com/politics/videocirculating-of-claudia-sheinbaum-is-apparently-a-deepfake>, fecha de consulta: 24 de juniode 2025.

Ministério da Ciência, Tecnologia e Inovação (MCTI), Brazil, "IA para o Bem de Todos: Proposta de Plano Brasileiro de Inteligência Artificial 2024-2028" [IA para el bien de todos: propuesta de plan brasileño de inteligencia artificial 2024-2028], 29 de julio de 2025, <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2024/07/plano-brasileiro-de-ia-tera-supercomputador-e-investimento-de-r-23-bilhoes-em-quatro-anos/ia_para_o_bem_de_todos.pdf/view>, fecha de consulta: 17 de junio de 2025.

Mont'Alverne, C., Arguedas, A. R., Toff, B. y Badrinathan, S., "Trust in the media in Latin America: The same old story?" [Confianza en los medios en América Latina: ¿la misma historia de siempre?], LSE Latin America and Caribbean Blog, 2 de febrero

de 2021, <https://blogs.lse.ac.uk/latamcaribbean/2021/02/02/trust-in-the-media-in-latin-america-the-same-old-story>, fecha de consulta: 2 de junio de 2025.

Mozilla Foundation, "YouTube Regrets: A Crowdsourced Investigation into YouTube's Recommendation Algorithm" [YouTube se arrepiente: una investigación colaborativa sobre el algoritmo de recomendaciones de YouTube], julio de 2021, <https://foundation.mozilla.org/en/youtube/findings>, fecha de consulta: 2 de junio de 2025.

Murphy, H., "The trouble with deepfakes—liar's dividend" [El problema de los deepfakes: el dividendo del mentiroso], Financial Times Tech Tonic Podcast, 22 de agosto de 2024, <https://www.ft.com/content/7f22ce59-1c6c-4d84-bca8-dc539992e286>, fecha de consulta: 17 de junio de 2025.

Naciones Unidas México, "Inteligencia artificial centrada en los pueblos indígenas: Perspectivas desde América Latina", 14 de diciembre de 2023, <https://mexico.un.org/es/255872-inteligencia-artificial-centrada-en-los-pueblos-ind%C3%ADgenas-perspectivas-desde-am%C3%A9rica-latina>, fecha de consulta: 2 de junio de 2025.

Narayanan, A., "Understanding social media recommendation algorithms" [Comprender los algoritmos de recomendación de las redes sociales], Knight First Amendment Institute, Columbia University, 9 de marzo de 2023, <https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms>, fecha de consulta: 2 de junio de 2025.

Nicholas, G. y Bhatia, A., "Lost in translation: Large language models in non-English content analysis" [Perdidos en la traducción: modelos extensos de lenguaje en el análisis de contenido no inglés], 2023, <https://doi.org/10.48550/arXiv.2306.07377>.

Office of the United Nations High Commissioner for Human Rights (OHCHR), "General Comment No. 25: The right to participate in public affairs, voting rights and the right of equal access to public service (Art. 25):. 12/07/1996 CCPR/C/21/Rev.1/Add.7, General Comment No. 25" [Observación General Nº 25: El derecho a participar en los asuntos públicos, el derecho de voto y el derecho de acceso, en condiciones de igualdad, a las funciones públicas (art. 25): 12/07/1996 CCPR/C/21/Rev.1/Add.7, Observación General Nº 25], <https://www.equalrightstrust.org/ertdocumentbank/general%20comment%2025.pdf>, fecha de consulta: 2 de junio de 2025.

Organization for Security and Co-operation in Europe (OSCE), Office of the Representative on Freedom of the Media, *Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual* [Enfoque en la inteligencia artificial y la libertad de expresión: un manual de políticas] (Vienna: OSCE, 2021), <https://www.osce.org/files/f/documents/8/f/510332_1.pdf>, fecha de consulta: 2 de junio de 2025.

Ormerod, A. G., "Latin Americans' social feeds are being flooded with AI-generated ads" [Las redes sociales de los latinoamericanos se están inundando de anuncios generados por IA], Rest of the World, 19 de julio de 2023, <https://restofworld.org/2023/newsletter-latin-america-generative-ai-ads>, fecha de consulta: 17 de junio de 2025.

Oversight Board, "Content Moderation in a New Era for AI and Automation" [Moderación de contenido en una nueva era para la IA y la automatización], septiembre 2024, <https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation>, fecha de consulta: 2 de junio de 2025.

Panoptykon Fundacja, "Fixing recommender systems. From identification of risk factors to meaningful transparency and mitigation" [Mejora de los sistemas

de recomendación. Desde la identificación de factores de riesgo hasta la transparencia y la mitigación significativas], Briefing Note, 25 agosto 2023, <https://en.panoptykon.org/fixing-recommender-systems-identification-risk-factors-meaningful-transparency-and-mitigation>, fecha de consulta: 17 de junio de 2025.

Piña, G., "Cinco casos de 'DeepFake porn' en México. La ley es insuficiente: Olimpia Coral", EMEEQUIS, 29 de junio de 2023, <https://m-x.com.mx/entrevistas/cinco-casos-de-deepfake-porn-en-mexico-la-ley-es-insuficiente-olimpia-coral>, fecha de consulta: 2 de junio de 2025.

Poder360, "Bolsonaro sugere uso de IA para questionar Alexandre de Moraes", 22 de febrero de 2025, <https://www.poder360.com.br/poder-justica/bolsonaro-sugere-uso-de-ia-para-questionar-alexandre-de-moraes>, fecha de consulta: 2 de junio de 2025.

Rangel, L., "INE registró 205 quejas de violencia política de género en este proceso electoral", Animal Político, 29 de junio de 2024, <https://animalpolitico.com/verificacion-de-hechos/te-explico/ine-violencia-politica-genero-elecciones>, fecha de consulta: 17 de junio de 2025.

Resk, F., "Tabata usa deep fake para expor Nunes como Ken de Barbie e refaz post" [Tábata utiliza un profundo fake para exhibir a Nunes como el Ken de Barbie y rehace su post], Metropoles, 5 de abril de 2024, <https://www.metropoles.com/sao-paulo/tabata-deep-fake-nunes-ken-barbie>, fecha de consulta: 17 de junio de 2025.

Rid, T., *Active Measures: The Secret History of Disinformation and Political Warfare* [Medidas activas: La historia secreta de la desinformación y la guerra política] (Nueva York: Farrar, Straus and Giroux, 2020).

—, "The lies Russia tells itself: The country's propagandists target the West—but mislead the Kremlin, too" [Las mentiras que Rusia se cuenta a sí misma: los propagandistas del país atacan a Occidente, pero también engañan al Kremlin], *Foreign Affairs*, 30 de septiembre de 2024, <https://www.foreignaffairs.com/united-states/lies-russia-tells-itself>, fecha de consulta: 2 de junio de 2025.

Ríos Tobar, M., *Violencia política de género en la esfera digital en América Latina* (Estocolmo: IDEA Internacional, 2024), <https://doi.org/10.31752/idea.2024.93>.

Rodríguez, A., "Acusado un universitario de alterar con inteligencia artificial miles de imágenes de alumnas para venderlas como pornografía", *El País*, 13 de octubre de 2023, <https://elpais.com/mexico/2023-10-14/acusado-un-universitario-de-alterar-con-inteligencia-artificial-miles-de-imagenes-de-alumnas-para-venderlas-como-pornografia.html>, fecha de consulta: 2 de junio de 2025.

Shahid, F., "Colonialism in content moderation research: The struggles of Scholars in Majority World" [El colonialismo en la investigación sobre moderación de contenidos: las luchas de los académicos en el mundo mayoritario], Center for Democracy & Technology, 26 de agosto de 2024, <https://cdt.org/insights/colonialism-in-content-moderation-research-the-struggles-of-scholars-in-the-majority-world>, fecha de consulta: 18 de junio de 2025.

Simon, F. M., Altay, S. y Mercier, H., "Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown" [¿Desinformación recargada? Los temores sobre el impacto de la IA generativa en la desinformación son exagerados], *HKS Misinformation Review*, 18 de octubre de 2023, <https://doi.org/10.37016/mr-2020-127>.

Soliman, W. y Rinta-Kahila, T., "Unethical but not illegal! A critical look at two-sided disinformation platforms: Justifications, critique, and a way forward" [¡Poco

ético, pero no ilegal! Una mirada crítica a las plataformas de desinformación de dos caras: Justificaciones, crítica y un camino a seguir], *Journal of Information Technology*, 39/3 (2024), págs. 441-476, <https://doi.org/10.1177/02683962231181145>.

Souza, C. A., "A Blueprint for Digital Propaganda in the Majority World: Online Coordinated Attacks, Narratives About Content Moderation and Challenges to Electoral Integrity in Brazil" [Un modelo para la propaganda digital en el mundo mayoritario: ataques coordinados en línea, narrativas sobre moderación de contenido y desafíos a la integridad electoral en Brasil], Yale Law School, Information Society Project, Majority World Initiative, julio de 2024, <https://law.yale.edu/sites/default/files/area/center/isp/documents/mwi-carlos-affonso-souza_2024-08-01_re-fin.pdf>, fecha de consulta: 2 de junio de 2025.

Souza, L. y Varon, J., "Violencia Política de Género en Internet", Policy Paper, Al Sur, 2019, <https://www.alsur.lat/sites/default/files/2021-07/Violencia Pol%C3%ADtica de G%C3%A9nero en Internet ES.pdf>, fecha de consulta: 2 de junio de 2025.

Törnberg, P. y Chueri, J., "When do parties lie? Misinformation and radical-right populism across 26 countries" [¿Cuándo mienten los partidos? Desinformación y populismo de extrema derecha en 26 países], *The International Journal of Press/Politics*, 0/0 (2025), <https://doi.org/10.1177/19401612241311886>.

Tribunal Regional Eleitoral–BA, "Chatbot Maia passa a oferecer informações sobre Registro de Candidaturas" [Chatbot Maia ahora ofrece información sobre el Registro de Candidatos], 23 de octubre de 2024, <https://www.tre-ba.jus.br/comunicacao/noticias/2023/Outubro/chatbot-maia-passa-a-oferecer-informacoes-sobre-registro-de-candidaturas>, fecha de consulta: 18 de junio de 2025.

Tribunal Superior Eleitoral (TSE), Resolução nº 23.732, de 27 de fevereiro de 2024 [Resolución Nº 23.732], Brasília, 27 de febrero de 2024, <https://www.tse.jus.br/legislacao/compilada/res/2024/resolucao-no-23-732-de-27-de-fevereiro-de-2024>, fecha de consulta: 2 de junio de 2025.

Udupa, S., Maronikolakis, A. y Wisiorek. A., "Ethical scaling of content moderation: Extreme speech and the insignificance of artificial intelligence" [Escala ética de la moderación de contenidos: El discurso extremo y la insignificancia de la inteligencia artificial], *Big Data & Society*, 10/1 (2023), <https://doi.org/10.1177/20539517231172424>.

United Nations, Promotion and Protection of the Right to Freedom of Opinion and Expression [Promoción y protección del derecho a la libertad de opinión y expresión], A/73/348, 29 de agosto de 2018, <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N18/270/42/PDF/N1827042.pdf>, fecha de consulta: 2 de junio de 2025.

—, *Governing AI for Humanity: Final Report* [Gobernanza de la IA para la humanidad: Informe final] (Nueva York, UN, 2024), <https://digitallibrary.un.org/record/4062495?v=pdf>, fecha de consulta: 19 de junio de 2025.

United Nations Educational, Scientific and Cultural Organization (UNESCO), *Artificial Intelligence and Democracy* [Inteligencia artificial y democracia] (Paris, UNESCO, 2024) <https://www.unesco.org/en/articles/artificial-intelligence-and-democracy>, fecha de consulta: 2 de junio de 2025.

Wood, B., "Could Bolivia's current politics be fueling Indigenous discrimination?" [¿Podría la actual política de Bolivia estar alimentando la discriminación indígena?], PBS NewsHour, 25 de noviembre de 2019, <https://www.pbs.org/newshour/world/

could-bolivias-current-politics-be-fueling-indigenous-discrimination>, fecha de consulta: 2 de junio de 2025.

Zanatta, R. A. F. y Rielli, M., "The artificial intelligence legislation in Brazil: Technical analysis of the text to be voted on in the federal senate plenary" [La legislación sobre inteligencia artificial en Brasil: análisis técnico del texto que será votado en el plenario del Senado federal], DataPrivacy Brazil Research, 12 de diciembre de 2024. <https://www.dataprivacybr.org/en/the-artificial-intelligence-legislation-in-brazil-technical-analysis-of-the-text-to-be-voted-on-in-the-federal-senate-plenary>, fecha de consulta: 19 de junio de 2025.

Sobre los autores

Sebastián Becker Castellaro es Oficial Asociado del Programa de Digitalización y Democracia de IDEA Internacional. Su trabajo se centra en el impacto de la tecnología en las democracias de todo el mundo, proporcionando investigación, análisis de políticas y apoyo a proyectos. Cuenta con una Maestría en Derecho Internacional de la Université Libre de Bruxelles y una Maestría en Derecho Público de la Universidad de Chile.

Mariana Carvalho es investigadora y profesora de FGV Comunicação. Es doctora en Comunicación por la Universidad Estatal de Río de Janeiro (UERJ) y licenciada en Periodismo por la Universidad Federal de Espírito Santo (Ufes). Su investigación se centra en las disputas narrativas y el papel de las emociones en el debate político en las plataformas digitales.

Alberto Fernández Gibaja es el responsable del Programa de Digitalización y Democracia de IDEA Internacional. Supervisa el trabajo del instituto en los aspectos que se encuentran en la intersección de la tecnología y la democracia. Tiene un máster en Estudios Internacionales por la Universidad de Uppsala y una licenciatura por la Universidad Complutense de Madrid.

Amaro Grassi es vicedirector de la Escuela de Comunicación, Medios e Información de la Fundação Getulio Vargas (FGV Comunicação). Es doctor en Ciencias Políticas y máster en Sociología por el Instituto de Estudios Sociales y Políticos de la Universidad del Estado de Río de Janeiro (IESP-UERJ).

Cecilia Hammar es Asistente de Programa del Programa de Digitalización y Democracia de IDEA Internacional. Lleva a cabo investigaciones y apoyo a la coordinación de proyectos, principalmente relacionados con las implicaciones sociales de la IA. Actualmente cursa una M.Sc. en Ciencias Políticas en la Universidad de Uppsala.

Juliane Müller es Oficial Asociada de Programas en el Programa de Digitalización y Democracia de IDEA Internacional. Se especializa en las implicaciones democráticas de las tecnologías emergentes, en particular la inteligencia artificial, con un enfoque en los derechos humanos y la integridad electoral. Su trabajo se centra en la iniciativa global de desarrollo de capacidades de IA del Programa para los organismos electorales e incluye investigación y análisis de políticas. Es licenciada en Derecho Internacional por la Universidad de Edimburgo y licenciada en Derecho por la Universidad de Mannheim.

Laura Pereira tiene una maestría en ciencias sociales por la Universidade Estadual Paulista (UNESP) y trabaja como líder de proyecto en FGV Comunicação. Su trabajo se centra en la intersección de Internet, la comunicación y la política, con experiencia en investigación y análisis de

políticas relacionadas con la regulación de la tecnología y los impactos políticos del consumo contemporáneo de medios.

Víctor Piaia es doctor en Sociología por el Instituto de Estudios Sociales y Políticos (IESP-UERJ). Es profesor de FGV Comunicação y líder del Grupo de Estudios en Sociología Política y Transformaciones Digitales. Estudia los efectos políticos de la transformación en la comunicación diaria, centrándose en las plataformas de redes sociales y las aplicaciones de mensajería.

Marco Aurélio Ruediger es doctor en Sociología y máster en Análisis y Gestión de Políticas. Actualmente es Director de la Escuela de Comunicación, Medios e Información de la Fundação Getulio Vargas (FGV Comunicação).

Acerca de IDEA Internacional

El Instituto Internacional para la Democracia y la Asistencia Electoral (IDEA Internacional) es una organización intergubernamental con 35 estados miembros, fundada en 1995 con el mandato de apoyar la democracia sostenible en todo el mundo.

QUÉ HACEMOS

Desarrollamos investigaciones favorables a las políticas y análisis relacionados con elecciones, parlamentos, constituciones, digitalización, cambio climático, inclusión y representación política, todo ello bajo el paraguas de los Objetivos de Desarrollo Sostenible de la ONU. Evaluamos el desempeño de las democracias a través de nuestro singular Índices del estado global de la democracia y el Observador de la democracia.

Ofrecemos desarrollo de capacidades y asesoramiento experto a los actores democráticos incluyendo a los gobiernos, los parlamentos, los funcionarios electorales y la sociedad civil. Desarrollamos herramientas y publicamos bases de datos, libros, y guías en varios idiomas sobre temas que van desde la participación electoral hasta las cuotas de género.

Acercamos a los actores estatales y no estatales para establecer diálogos y compartir aprendizajes. Nos hemos consagrado con promover y proteger la democracia mundial.

DÓNDE ACTÚA IDEA INTERNACIONAL

Nuestra sede se encuentra en Estocolmo. Contamos con oficinas regionales y nacionales en África, Asia-Pacífico, Europa y América Latina y el Caribe. IDEA Internacional es Observador Permanente de la ONU y está acreditada como institución en la Unión Europea.

NUESTRAS PUBLICACIONES Y BASES DE DATOS

Tenemos un catálogo con más de 1.000 publicaciones y más de 25 bases de datos en nuestro sitio web. La mayoría de nuestras publicaciones se pueden descargar de forma gratuita.

<https://www.idea.int>

FGV Comunicação

Fundada en 2022, la Escuela de Comunicación, Medios e Información (FGV Comunicação) es pionera en la formación de una nueva generación de profesionales capacitados para trabajar en los sectores de comunicación estratégica, medios de comunicación e información en instituciones públicas, empresas privadas y organizaciones de la sociedad civil. La Escuela combina la tradición académica de la FGV con la amplia experiencia de la FGV DAPP en la aplicación de la tecnología y ciencia de datos para impulsar innovación y soluciones en el campo de la comunicación.

FGV Comunicação cuenta con un distinguido cuerpo docente de Brasil y del extranjero, una metodología de enseñanza reconocida internacionalmente, un plan de estudios innovador, laboratorios de última generación y valiosas oportunidades de networking y desarrollo profesional para los estudiantes.

<https://comunicacao.fgv.br/>

IDEA Internacional
Strömsborg
SE-103 34 Estocolmo
SUECIA
+46 8 698 37 00
info@idea.int
www.idea.int

Este documento de política examina el impacto de la inteligencia artificial (IA) en la integridad de la información en contextos electorales en toda América Latina. Basándose en estudios de caso de las elecciones de 2024 en Brasil y México, analiza cómo se utilizan las herramientas de IA para crear, seleccionar y moderar contenidos políticos en línea, prestando especial atención a sus efectos en grupos marginados como las mujeres, las comunidades indígenas y las personas LGBTQIA+. Si bien el uso de la IA generativa en las elecciones sigue siendo limitado, su potencial para socavar el discurso democrático y amplificar el contenido dañino exige atención urgente.

Para proteger los procesos democráticos, este documento de política pide un marco regulatorio sólido que incluya obligaciones de transparencia para las plataformas, salvaguardas contra la violencia de género facilitada por la IA y estrategias de gobernanza inclusivas. Recomienda una supervisión más estricta del uso de la IA en las elecciones, sistemas de procedencia para el contenido sintético y una auditoría independiente de los algoritmos de curación de contenidos. A medida que la IA evoluciona, también debe hacerlo nuestro compromiso con la integridad electoral y la protección de todas las voces en la esfera democrática.

ISBN: 978-91-7671-976-3 (versión en pdf)