

LA DEMOCRACIA EN LA ERA DE LA INTELIGENCIA ARTIFICIAL GENERATIVA:

¿CÓMO AFRONTAR LOS RIESGOS Y APROVECHAR LAS OPORTUNIDADES?

**La democracia en la era de la inteligencia artificial (IA) generativa:
¿Cómo afrontar los riesgos y aprovechar las oportunidades?**

Copyright © 2024 International Republican Institute.

Reservados todos los derechos.

Declaración de permiso: Ninguna parte de este trabajo puede reproducirse en ninguna forma ni por ningún medio, electrónico o mecánico, incluyendo fotocopias, grabaciones o cualquier sistema de almacenamiento y recuperación de información sin el permiso por escrito del International Republican Institute.

Las solicitudes de permiso deben incluir la siguiente información:

- El título del documento para el que se solicita permiso para copiar material.
- Una descripción del material para el que se solicita permiso para copiar.
- El propósito para el que se utilizará el material copiado y la forma en que se utilizará.
- Su nombre, cargo, nombre de la empresa u organización, número de teléfono, dirección de correo electrónico y dirección postal.

Por favor enviar todas las solicitudes a:

Atención: Communications Department International Republican Institute
1225 Eye Street NW, Suite 800
Washington, DC 20005
info@iri.org

ACERCA DE NOSOTROS

THE INTERNATIONAL REPUBLICAN INSTITUTE (IRI)

El IRI es una de las principales organizaciones internacionales de desarrollo de la democracia del mundo. Este instituto no partidista y no gubernamental ha apoyado a organizaciones de la sociedad civil, periodistas, gobiernos democráticos y otros actores democráticos en más de 100 países desde 1983 (África, Asia, Eurasia, Europa, América Latina y el Caribe, Oriente Medio y el Norte de África) y tiene presencia en más de 80 países en todo el mundo. La Práctica de Tecnología y Democracia del IRI es perceptible en todas las regiones del mundo para ayudar a los actores de base a utilizar la tecnología para promover los principios democráticos, resolver problemas públicos y apoyar una gobernanza más sólida.

PRÁCTICA DE TECNOLOGÍA Y DEMOCRACIA DEL IRI

La programación de Tecnología y Democracia del IRI tiene como objetivo hacer que la revolución digital trabaje a favor de la democracia y no en contra de ella, específicamente:

1. Garantizando que la revolución digital promueva la democracia y la libertad y supere a los autoritarios digitales;
2. Construyendo una democracia del siglo XXI galvanizando los beneficios de la tecnología para fortalecer la democracia; y
3. Aumentando la resiliencia de los actores democráticos ante la represión digital.

El Departamento de Tecnología y Democracia del IRI ha llevado a cabo programas sobre democracia digital, tecnología cívica y gubernamental, libertad en Internet, tecnologías emergentes e integridad de la información en todas las regiones del mundo. El IRI tiene experiencia en apoyar a organizaciones de la sociedad civil (OSC) y otros socios de la democracia, como gobiernos, partidos políticos, activistas y otros, en la incorporación de herramientas de libertad en Internet, y en el monitoreo, la lucha y la creación de resiliencia ante amenazas a la integridad de la información en entornos represivos, así como en el desarrollo de programas de tecnología para una buena gobernanza. Estos programas han incluido el apoyo a los actores para implementar tecnologías digitales de manera que promuevan la transparencia, la inclusión, la rendición de cuentas y la capacidad de respuesta, al tiempo que se adaptan a un panorama de amenazas cambiante, incluidas las tecnologías emergentes.

AGRADECIMIENTOS

La elaboración de este informe técnico fue posible gracias a la redacción y la investigación de Hui Hui Ooi, Aaron Spitler y Amanda Zink, con el amplio apoyo de Justin Sherman y Devan Desai de Global Cyber Strategies. Agradecemos a Jared Ford por su revisión, apoyo editorial y orientación estratégica durante el desarrollo y la finalización del informe. También agradecemos a los numerosos miembros del personal del IRI que brindaron comentarios invaluableles durante el proceso de redacción.

- Nicklas Badum, Democracy X
- Laure de Roucy-Rochegonde, Institut français des relations internationales (IFRI)
- Renee DiResta, Independent
- Konur Metehan Durmaz, Social Media Exchange (SMEX)
- Adam Fivenson, National Endowment for Democracy (NED)
- Allie Funk, Freedom House
- Arzak Khan, Innovation Center for Artificial Intelligence and Human Rights
- Agita Pasaribu, Bullyid Indonesia
- Cristina Martinez Pinto, PIT Policy Lab
- Lisa Reppell, International Foundation for Electoral Systems (IFES)
- Max Rizzuto, Atlantic Council's Digital Forensic Research Lab (DFRLab)
- Priscilla Ruiz, Article 19 Mexico and Central America Regional Office
- Frei Sangil, Layertech Lab
- Maurice Sayinzoga, National Democratic Institute (NDI)

LISTA DE ACRÓNIMOS

AFP: Agence France-Presse (Agencia de prensa francesa)

CCP: Chinese Community Party (Partido Comunista Chino)

OSC: Organización de la sociedad civil

C2PA: Steering Committee of the Coalition for Content Provenance and Authenticity (Comité Directivo de la Coalición para la Procedencia y Autenticidad del Contenido)

DC: Distrito de Columbia

DFRLab: Digital Forensic Research Lab at the Atlantic Council (Laboratorio de investigación forense digital del Atlantic Council)

CEI: Comisión Electoral de la India

UE: Unión Europea

AMIA: Alianza mundial sobre inteligencia artificial

GRAIL: Laboratorio de gobernanza e inteligencia artificial responsable de la Universidad de Purdue

IFES: Fundación Internacional para Sistemas Electorales

IFRI: Instituto Francés de Relaciones Internacionales

IONGI: Organización no gubernamental internacional

PRI Plan de respuesta a incidentes

US: Estados Unidos

USAID: United States Agency for International Development (Agencia de los Estados Unidos para el Desarrollo Internacional)

AGNU: Asamblea General de las Naciones Unidas

TI: Tecnología de la información

LGBTQI+: lesbianas, gays, bisexuales, transgénero, queer, intersexuales y más

LLM: Large language model (modelo de lenguaje amplio)

MIT: Massachusetts Institute of Technology (Instituto Tecnológico de Massachusetts)

ML: Machine learning (Aprendizaje automático)

NED: National Endowment for Democracy (Fundación Nacional para la Democracia)

PNL: Procesamiento natural del lenguaje

ONG: Organización no gubernamental

IRI: International Republican Institute (Instituto Republicano Internacional)

OEDE: Organización para la Cooperación y el Desarrollo Económico

RPC: República popular de China

RAIN: Responsible AI Network (Africa)

SMEX: Social Media Exchange

TFGBV: Tecnología facilitadora de la violencia de género

PNUD: Programa de las Naciones Unidas para el Desarrollo

UNESCO: United Nations Educational, Scientific and Cultural Organization (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura)

TABLA DE CONTENIDO

Introducción.....	1
Hallazgos clave.....	4
Introducción informativa: Entender la IA generativa.....	7
Riesgos de la IA generativa para la democracia.....	9
Oportunidades de la IA generativa para la democracia.....	20
Recomendaciones.....	27
Mirando hacia el futuro: ¿Qué anticipar?.....	38
Apéndice.....	40

INTRODUCCIÓN: ¿POR QUÉ ESCRIBIMOS ESTO?

Las sociedades democráticas no son ajenas al cambio. A lo largo de la historia, las democracias han evolucionado en respuesta a un conjunto diverso de factores sociales, políticos y económicos una y otra vez. Aun así, hay momentos en la historia en los que las sociedades democráticas se enfrentan a un ritmo significativo de cambio y a una presión para que se produzca. El auge de la inteligencia artificial generativa (IAGen) a finales de 2022 parece haber inaugurado uno de esos momentos, prometiendo una transformación en el funcionamiento de la sociedad y el gobierno que muchos han comparado con las transformaciones provocadas por la llegada de Internet.¹ Como ha sucedido con muchos avances tecnológicos del pasado, los gobiernos y la sociedad civil pueden tener dificultades para comprender, anticipar y minimizar los daños y maximizar los beneficios de las innovaciones digitales en las sociedades democráticas. Para evitar que se repitan los errores del pasado, la Práctica de Tecnología y Democracia del Instituto Republicano Internacional (IRI) lanzó un programa en el verano de 2023 para responder a las demandas de los socios sobre la preparación para los impactos actuales y potenciales causados por la IAGen).²

Muchos de los socios del IRI, en particular los gobiernos y los actores de la sociedad civil, tienen dificultades para seguir el ritmo de las innovaciones tecnológicas, a menudo no aprovechan las nuevas herramientas en beneficio de la democracia y tienen dificultades para promover o regular de manera proactiva sus efectos nocivos. Esto es especialmente preocupante, ya que estos dos grupos de interesados son clave para el funcionamiento eficaz de las democracias. Los socios del IRI (en este caso, la sociedad civil y los encargados de la toma de decisiones, incluidos los funcionarios electos y los encargados de formular políticas) han reconocido cada vez más la necesidad de mejorar sus competencias en este tema para garantizar que las democracias sigan siendo ágiles, modernas y eficaces en la era digital.

Esta mejora intencional de las competencias es especialmente importante para garantizar que los actores del Sur Global no se queden atrás. Este informe técnico pretende ayudar a llenar los vacíos de conocimiento y ayudar a los socios del IRI a comprender mejor esta innovación, proporcionando un análisis de los riesgos, desafíos y oportunidades de la IAGen y la democracia para los actores gubernamentales y de la sociedad civil de menor capacidad en el Norte Global y el Sur Global. El informe técnico proporciona una descripción general escrita para presentar y definir la IAGen, incluido un análisis de las principales amenazas y riesgos que plantea la IAGen para la democracia, y un resumen de las oportunidades para que la sociedad civil y los responsables de las políticas la aprovechen para el bien común.

Este informe técnico tiene como objetivo ayudar a llenar los vacíos de conocimiento y ayudar a los socios del IRI a comprender mejor la IA generativa, brindando un análisis de los riesgos, desafíos y oportunidades para la IA generativa y la democracia para los actores de la sociedad civil y los gobiernos de menor capacidad en todo el mundo.

El informe técnico también incluye un capítulo sobre recomendaciones para la sociedad civil, los responsables de las políticas y la industria para garantizar que IAGen trabaje a favor de la democracia, y no en contra de ella. El capítulo final del informe técnico se centra en los impactos a corto plazo de IAGen, ofreciendo a los lectores la oportunidad de considerar posibles impactos futuros. Es importante tener en cuenta que los hallazgos del informe técnico están generalizados para que sean ampliamente aplicables a diversos contextos globales, y que las amenazas, oportunidades y recomendaciones pueden variar según estos factores.

¹ Bajarin, Tim. "The practical impact of AI for the Masses." Forbes, 28 Nov. 2023, www.forbes.com/sites/timbajarin/2023/11/28/the-practical-impact-of-ai-for-the-masses/?sh=5217fc8f675b.

² Hijjawi, Faisal. "AI's Regulatory Challenge: Avoiding the Pitfalls of Past Mistakes." Science and Technology Law Review, 18 Apr. 2023, journals.library.columbia.edu/index.php/stlr/blog/view/583.

Este informe técnico es la culminación de un extenso análisis del panorama y una serie de entrevistas, seguido por un grupo de trabajo de seis meses de duración que convocó regularmente a más de 20 expertos, incluidos representantes de la sociedad civil, el mundo académico, la industria y las políticas, para discutir los impactos existentes y potenciales que la IAGen tendrá en las democracias, con un enfoque particular en las tendencias en el Sur Global.³ El informe técnico es una síntesis de los debates y hallazgos del grupo de trabajo, destinado a brindar orientación a la sociedad civil y a los responsables de las políticas sobre cómo maximizar los beneficios y minimizar los daños que la IAGen representa para la democracia.

³ "Generative Artificial Intelligence and Democracy Working Group." International Republican Institute, www.iri.org/what-we-do/programs/generative-artificial-intelligence-and-democracy-working-group/.

PARA QUIÉN FUE ESCRITO ESTE INFORME

Los riesgos, oportunidades y recomendaciones incluidos en este informe técnico son relevantes para cualquier actor preocupado por el futuro de la democracia. Para algunos lectores, puede ser un área de interés obvia, pero para otros puede parecer alejada de su trabajo actual o remota de los desafíos que enfrentan.

La industria, las organizaciones de la sociedad civil (OSC) y los gobiernos, todos actores clave en las sociedades democráticas, han comenzado a utilizar estas herramientas, por lo que es necesario que al menos estén familiarizados con los conceptos básicos. A continuación, se muestra una muestra de los tipos de grupos que podrían encontrar beneficioso este informe técnico; incluso si su organización o gobierno actualmente no interactúa con la IAGen ni lo utiliza, estas herramientas eventualmente comenzarán a aparecer en sus áreas de trabajo.

- OSC de defensa de derechos
- OSC de participación ciudadana
- Responsables de las políticas
- Funcionarios electos
- OSC de sensibilización y educación
- Grupos de verificación de datos y/o medios de comunicación
- Empresas tecnológicas

Considere el siguiente ejemplo: usted trabaja en una OSC centrada en la participación ciudadana que no ha decidido si utilizará o interactuará intencionalmente con herramientas de la IAGen. Sin embargo, en los próximos meses o incluso semanas, su gobierno lanzará un chatbot virtual que aprovecha IAGen para abordar las inquietudes, quejas y sugerencias de los ciudadanos. O tal vez una OSC asociada comenzará a utilizar herramientas IAGen para traducir de manera rápida y eficiente guías que explican cómo las personas pueden acceder a los servicios gubernamentales a varios idiomas. En cualquier caso, será importante, como mínimo, comprender estas herramientas, los riesgos que presentan y las prácticas de datos que se emplearán para educar a los ciudadanos y otros beneficiarios. Es probable que estas situaciones surjan cada vez más, por lo que es importante comprender este tema incluso si su organización no planea incorporar estas herramientas.



HALLAZGOS CLAVE

Tras una investigación documental, consultas con expertos y debates en grupos de trabajo sobre IAGen y sus impactos, en particular en todo el Sur Global, el IRI ha sintetizado los principales riesgos, oportunidades y recomendaciones aplicables a diversos actores democráticos. Estos hallazgos deberían tenerse en cuenta para ayudar a orientar el trabajo de la sociedad civil, los responsables de la toma de decisiones y los representantes de la industria que se esfuerzan por proteger la democracia y los principios democráticos a nivel mundial.



Riesgos clave

Mayor velocidad y menor costo de producción de desinformación

Exacerbación del discurso de odio, el acoso y la violencia de género facilitada por la tecnología (VGFT)

Reducción general de la confianza en el entorno de la información que puede conducir a un profundo escepticismo y una menor participación en los procesos e instituciones democráticos

Ampliación de las brechas digitales (especialmente con impacto en el Sur Global)

La falta de rendición de cuentas en el desarrollo y uso de IAGen puede reducir la confianza en la industria y el gobierno

Mayor preocupación por la privacidad, por ejemplo, qué datos se utilizan para desarrollar IAGen (y cómo se accedió a ellos), uso de GenAI para vigilancia y manipulación

Concentrarse demasiado en las soluciones técnicas (por ejemplo, la procedencia del contenido) a las amenazas introducidas por la IAGen puede desviar la atención de enfoques igualmente importantes que abordan el contexto político, regulatorio y social en el que existen las amenazas relacionadas con IAGen (por ejemplo, campañas de educación pública)



Oportunidades clave

Fortalecer las estrategias y los esfuerzos para proteger los entornos de información mediante una mejor identificación, etiquetado y desacreditación de la información errónea o desinformación

Potenciar la toma de decisiones de manera ágil y eficiente sintetizando documentos complejos, procesando grandes cantidades de datos para fundamentar la formulación de políticas y más

Mejorar la capacidad organizativa de los gobiernos y la sociedad civil que enfrentan limitaciones de capacidad

Mejorar la transparencia institucional aumentando la accesibilidad de los datos gubernamentales, la prestación de servicios, potenciando chatbots accesibles y más

Aprovechar la IAGen para respaldar la ampliación de los esfuerzos de educación cívica para proporcionar de manera más directa a las personas información clave sobre gobernanza y educación cívica

Crear ciberdefensas adaptables para mejorar las contramedidas defensivas y agilizar el trabajo necesario para defenderse de los ciberataques

Predecir con mayor precisión los resultados de las políticas para detectar tendencias emergentes, patrones y posibles escenarios para abordar de manera proactiva las situaciones

RECOMENDACIONES

A continuación, se presenta un conjunto de recomendaciones extraídas de las conversaciones de los grupos de trabajo para tres grupos de partes interesadas clave que son fundamentales para garantizar que la IAGen se aproveche democráticamente: la sociedad civil, el gobierno/los tomadores de decisiones y la industria. Cada grupo de partes interesadas debe tomar medidas para garantizar que la IAGen sea un beneficio (o, como mínimo, no un daño) para la democracia. Quizás aún más importante, estos grupos de partes interesadas deben colaborar entre sí para lograr un impacto real. Las recomendaciones que se encuentran a continuación están segmentadas para dirigirse específicamente a cada grupo de partes interesadas, y muchas de ellas hacen referencia a la colaboración de múltiples partes interesadas necesaria para abordar de manera eficaz los posibles daños que plantea la IAGen.



Sociedad civil

Consultar los marcos de uso ético de IAGen para desarrollar políticas internas de uso de IAGen

Unirse a coaliciones globales y abogar en esos foros por un mayor enfoque en IAGen y las regulaciones y la gobernanza de la IA

Desarrollar un plan de respuesta a incidentes (PRI) que incorpore cuestiones relacionadas con IAGen

Señalar los usos gubernamentales y del sector privado de IAGen que violan los derechos humanos

Apoyar los esfuerzos de educación social, como las campañas de educación cívica, así como la capacitación del personal interno



Gobierno/Tomadores de decisiones

Crear un sitio web de recursos públicos que describa de manera clara, comprensible y completa cómo se utiliza la IAGen en las agencias gubernamentales

Asegurarse de que los equipos de adquisiciones tengan personal multidisciplinario y un enfoque establecido para adquirir IAGen

Identificar oportunidades para integrar IAGen para mejorar la prestación de servicios y la accesibilidad de la información

Evaluar cómo IAGen se alinea con las leyes y regulaciones existentes

Consultar las pautas existentes para obtener información sobre cómo abordar las lagunas en las leyes y regulaciones actuales



Industria

Desarrollar capacidades entre los profesionales de la industria sobre las implicaciones legales, sociales y de derechos humanos de la IAGen

Involucrar a los responsables de las políticas en los procesos nacionales de elaboración de leyes sobre IA

Publicar más información sobre las prácticas de uso de IAGen y los estándares éticos en los sitios web de las empresas

Desarrollar productos y servicios de acuerdo con los principios de derechos humanos mientras se forman equipos multidisciplinarios y diversos

Proporcionar financiación sostenible a largo plazo para la sociedad civil y los socios académicos para garantizar que diversos actores puedan contribuir a las soluciones para minimizar los daños de la IAGen



Sociedad civil

Revisar las leyes y regulaciones de su país sobre los niveles de acceso a los datos relacionados con IAGen



Gobierno/Tomadores de decisiones

Involucrar las opiniones de los ciudadanos en todas las etapas de la integración de la IAGen

Involucrar a expertos externos y desarrollar la capacidad interna en la medida de lo posible



Industria

Crear oportunidades sostenidas de intercambio de conocimientos con la sociedad civil y el mundo académico para garantizar que diversas voces informen el desarrollo de herramientas, las políticas de uso y la creación de barreras de protección

Adaptar los esfuerzos de mitigación de daños a diversos entornos

INTRODUCCIÓN INFORMATIVA: ENTENDER LA IA GENERATIVA

Comprender la terminología

Para comprender la IAGen, que es un subconjunto de la inteligencia artificial (IA), es necesario comprender primero la definición general de IA (a menudo denominada IA tradicional o estrecha), que ha sido parte de los sistemas tecnológicos durante décadas. La Organización para la Cooperación y el Desarrollo Económicos (OCDE), que estableció los Principios de la OCDE sobre Inteligencia Artificial en 2019, define la IA como: "...un sistema basado en máquinas que puede, para un conjunto dado de objetivos definidos por humanos, hacer predicciones, recomendaciones o decisiones que influyen en entornos reales o virtuales".⁴ La IA se refiere a sistemas diseñados para responder a un conjunto particular de datos, impulsados por algoritmos,⁵ y se pueden utilizar para analizar datos y hacer predicciones sencillas. Estos sistemas han sido parte de la vida cotidiana durante años, como los algoritmos de IA que impulsan los motores de recomendación en Amazon, Netflix y Google. Estos sistemas de IA procesan entradas de datos, como información sobre los programas seleccionados previamente por los usuarios en el servicio de transmisión de Netflix o las palabras que los usuarios ingresan en el motor de búsqueda de Google, y luego brindan resultados sugeridos en respuesta a esas entradas. Estos sistemas de IA han sido programados para seguir reglas específicas para responder a las solicitudes, y no generan contenido nuevo.

Gran parte del entusiasmo actual en torno a la IA se centra específicamente en IAGen; el lanzamiento de ChatGPT a finales de 2022 puso a la IAGen en primer plano y la hizo accesible a los ciudadanos comunes sin experiencia en IAGen, a menudo de forma gratuita. IAGen, aunque se basa en un sistema similar basado en máquinas, también puede crear contenido nuevo en respuesta a indicaciones, incluidas imágenes, texto y contenido basado en audio. Sin embargo, en lugar de simplemente seguir un conjunto específico de reglas, los modelos IAGen se entrenan con un conjunto de datos y aprenden los patrones subyacentes a través de un proceso llamado aprendizaje automático (ML - Machine Learning, por sus siglas en inglés) para generar nuevos datos que reflejen el conjunto de entrenamiento. Esto también puede describirse como el cálculo de la probabilidad del siguiente token en una secuencia, como cuál podría ser la siguiente palabra en una oración. La IAGen se basa en grandes modelos de lenguaje (LLM - Large Language Models por sus siglas en inglés) además de otros tipos de modelos de aprendizaje automático. Las innovaciones recientes en IAGen incluyen chatbots interactivos y portales de búsqueda como ChatGPT de OpenAI, Bing de Microsoft y Gemini de Google, todos los cuales pueden generar resultados en respuesta a entradas de varios tipos. Esta funcionalidad incluye, por ejemplo, generar un esquema para un trabajo académico con instrucciones limitadas o crear una imagen de una ballena a pedido. Cuantos más datos reciba un sistema IAGen, más sólidos serán sus resultados.

Hay muchos términos relacionados con la IAGen que solo se han abordado brevemente en este manual. Un buen punto de partida es familiarizarse con algunos de los términos que surgen con frecuencia en las conversaciones sobre IA, como algoritmos, aprendizaje automático y LLM. Sugerimos revisar el Glosario de términos de IA de CNET para obtener una descripción general fácil de usar de las definiciones clave relacionadas con la comprensión de la IA y IAGen.⁶ Hay recursos más detallados que pueden ayudar a comprender cómo se interconectan estos términos, como la serie de blogs Generative AI for Beginners de Medium.⁷

⁴ Artificial Intelligence and Responsible Business Conduct. OECD, 2021, mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf.

⁵ Algorithms are step-by-step instructions or rules that guide AI to solve a specific problem or perform a particular task. For GenAI, that can include instructing GenAI to learn from data, recognize patterns, generate new content, and adapt over time.

⁶ Khan, Imad. "CHATGPT Glossary: 44 AI Terms That Everyone Should Know," CNET, 14 May 2024, www.cnet.com/tech/services-and-software/chatgpt-glossary-44-ai-terms-that-everyone-should-know/.

⁷ Gupta, Raja. "Generative AI for Beginners: Part 1 – Introduction to AI," Medium, 8 Feb 2024, <https://medium.com/@raja.gupta20/generative-ai-for-beginners-part-1-introduction-to-ai-eadb5a71f07d>. Archived Version, <https://web.archive.org/web/20240719150037/https://medium.com/@raja.gupta20/generative-ai-for-beginners-part-1-introduction-to-ai-eadb5a71f07d>.

Variables que determinan el uso actual

Antes de analizar los riesgos, amenazas y oportunidades que presenta la IAGen para la democracia, es importante señalar cómo se ha utilizado la IAGen hasta la fecha y en qué punto se encuentran las conversaciones actuales sobre la IAGen. Es especialmente importante considerar cómo pueden variar los impactos de la IAGen según el contexto político de un país. En el breve tiempo transcurrido desde que OpenAI lanzó ChatGPT, su uso se ha disparado en el ámbito académico, entre los ciudadanos comunes, la industria, la sociedad civil, el sector público y otros; más adelante en el informe técnico se analizarán algunos de estos casos de uso. Lo que es especialmente importante señalar, incluso para los lectores fuera del Norte Global, es que la forma en que se ha desarrollado el uso y el impacto de la IAGen en los distintos países varía según los entornos regulatorios, el acceso a Internet y la infraestructura, las normas éticas y los niveles de alfabetización digital, entre otros factores.

Las conversaciones y los análisis sobre los impactos de la IAGen suelen centrarse de forma desequilibrada en el Norte Global, donde la infraestructura, la privacidad de los datos y el acceso a Internet suelen ser más fiables, y donde se originan las normas y culturas incorporadas al desarrollo de estas herramientas. Se habla menos de cómo se utiliza o no la IAGen en el Sur Global, donde una infraestructura más débil, una Internet más lenta y una falta de regulación protectora pueden dar lugar a daños descomunales, incluida la exacerbación de la brecha digital. Rara vez se mencionan las realidades de las subdivisiones digitales, como las disparidades entre el acceso urbano y rural, las desigualdades de género y el acceso limitado para las comunidades marginadas. Marcos como la Evaluación del Ecosistema Digital de los Países de USAID y recursos como Civicspace.tech pueden ayudar a analizar los ecosistemas digitales y considerar los espacios cívicos para informar debates más matizados sobre los impactos de la IAGen, incluida la forma en que la IAGen puede exacerbar o resolver los desafíos existentes.

También es importante señalar la rápida aparición de regulaciones y marcos de IA que parecen muy diferentes según el gobierno, la entidad o el país. Por ejemplo, al momento de escribir este informe técnico, la Unión Europea (UE) publicó la Ley de IA de la UE,⁸ En el caso de la IA, la política de la Casa Blanca de los Estados Unidos (EE. UU.) se basa en un enfoque basado en el riesgo para gobernar el avance de la IA, mientras que el Proyecto de la Casa Blanca para una Ley de Derechos de la IA⁹ enfatiza principios rectores no vinculantes en lugar de reglas específicas basadas en riesgos. Ambos enfatizan el respeto por los derechos humanos, la transparencia y la gestión de riesgos. La República Popular China (RPC), por su parte, ha desarrollado las Medidas provisionales para la gestión de los servicios de IA generativa, un enfoque basado en reglas que es rígido y se centra en aplicaciones específicas de la tecnología.¹⁰ Las principales organizaciones no gubernamentales internacionales (ONGI) han elaborado marcos adicionales, como las Recomendaciones de la UNESCO sobre la ética de la IA,¹¹ La Resolución de la Asamblea General de las Naciones Unidas sobre IA,¹² y los Principios de IA de la OCDE,¹³ para nombrar solo algunos. Al considerar las implicaciones de la IA y la IAGen en un contexto particular, es importante recordar que los marcos o regulaciones existentes pueden variar ampliamente (o estar completamente ausentes) en dicho país, lo que altera el impacto general que puede tener la IAGen. El AI Watch: Global Regulatory Tracker¹⁴ puede ser una herramienta útil para ayudar a hacer un seguimiento de las políticas en evolución, así como el Índice Global sobre IA Responsable del Centro Global de Gobernanza de la IA.¹⁵

⁸ "The EU Artificial Intelligence Act." European Union, artificialintelligenceact.eu/. Accessed 26 June 2024.

⁹ "Blueprint for an AI Bill of Rights," The White House, Oct. 2022, www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

¹⁰ Regulatory and Legislation: China's Interim Measures for the Management of Generative Artificial Intelligence Services Officially Implemented. PwC, Aug. 2023, www.pwccn.com/en/tmt/interim-measures-for-generative-ai-services-implemented-aug2023.pdf.

¹¹ Recommendation on the Ethics of Artificial Intelligence, UNESCO, 2022, unesdoc.unesco.org/ark:/48223/pf0000381137.

¹² General Assembly Adopts Landmark Resolution on Artificial Intelligence," UN News, 21 Mar. 2024, news.un.org/en/story/2024/03/1147831.

¹³ "Artificial Intelligence," OECD.

¹⁴ "AI Watch: Global Regulatory Tracker." White & Case Insights, 13 May 2024, www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker.

¹⁵ "Main Findings." Global Index on Responsible AI, 13 June 2024, <https://global-index.ai/Takeaways>.

RIESGOS DE LA IA GENERATIVA PARA LA DEMOCRACIA

Desde el lanzamiento de ChatGPT a fines de 2022, seguido de herramientas como Claude¹⁶, Gemini¹⁷, y Meta AI¹⁸, se ha producido un aumento de la atención y el uso público de las herramientas de IAGen, lo que ha aumentado los riesgos que se plantean a la democracia, incluso para los países del Sur Global. Los riesgos relacionados con la IAGen se han presentado como daños reales que amenazan con empeorar. La difusión de deepfakes (contenido falsos), que se refieren a contenido audiovisual que ha sido manipulado de manera convincente para tergiversar las acciones de alguien, el uso indebido autoritario y los riesgos para la privacidad de los datos están exacerbando tendencias que pueden erosionar los pilares básicos de una democracia saludable; en concreto, el (mal) uso de la IAGen ha plagado los ecosistemas de información en línea, ha interferido en las elecciones, ha silenciado voces necesarias para el pluralismo democrático y ha reducido la confianza pública en las instituciones. A continuación, se presentan detalles adicionales de las posibles amenazas de la IAGen a la democracia.

Mayor velocidad y menor costo de producción de desinformación

os actores autoritarios en países como China y Rusia, así como los actores iliberales dentro de los países democráticos, pueden usar IAGen para acelerar la producción de desinformación a un menor costo. La credibilidad de este contenido varía. Aquellos con un ojo más perspicaz pueden analizar los movimientos de la boca en videos deepfakes, por ejemplo, para determinar la autenticidad de la información; pero no todos tendrán un sentido de detección tan agudo. A menudo, los grupos marginados, como las mujeres y las minorías étnicas, tienen niveles más bajos de alfabetización digital. Por lo tanto, existe el riesgo de que puedan verse desproporcionadamente afectados o influenciados por ese tipo de contenido. En algunos casos, los grupos no estatales emprendedores o los individuos que buscan ganancias financieras en el mercado de la desinformación aprovechan esta tecnología, que se alimenta del uso accesible y de bajo costo de la IAGen. Este riesgo es especialmente dañino en las elecciones, ya que la rápida proliferación de herramientas de IAGen significará que los votantes comunes se encontrarán con desinformación fácil de crear que puede ser cada vez más convincente a medida que estas herramientas se vuelvan más sofisticadas. Esto puede afectar la capacidad de los votantes para autenticar la información y tomar decisiones informadas, lo que podría resultar en una mayor apatía política y puede debilitar a los candidatos vulnerables, incluidas las mujeres. Si bien aún está por verse el impacto del contenido creado por IAGen en los resultados electorales y es complejo determinarlo, estos factores son preocupantes para los actores de la democracia en todo el mundo. Los países de todo el mundo, incluidos los del Sur Global, que ya luchan contra las amenazas a la integridad de la información en línea y carecen de los recursos adecuados para hacerlo, enfrentarán más desafíos a medida que las falsificaciones profundas invadan cada vez más el espacio de información en línea.



Caso de estudio

ESLOVAQUIA | DESINFORMACIÓN ELECTORAL GENERADA POR IA

En vísperas de las elecciones parlamentarias eslovacas de 2023, circuló en Facebook un clip de audio que supuestamente captaba una conversación entre Michal Šimečka, líder del partido liberal Eslovaquia Progresista, y Monika Tódová, periodista del periódico de centroderecha Denník N. En la grabación supuestamente aparecía Šimečka explicando a Tódová cómo pretendía robar la contienda a sus rivales, principalmente sobornando a votantes de la marginada comunidad gitana del país. Tanto el político como la periodista respondieron que el audio era completamente inventado, y un informe del departamento de verificación de datos de la Agence France-Presse (AFP) concluyó posteriormente que lo más probable es que el clip hubiera sido alterado mediante herramientas de inteligencia artificial.¹⁹

¹⁶ "Home." Claude, claude.ai. Accessed 26 June 2024.

¹⁷ "Home." Gemini, gemini.google.com. Accessed 26 June 2024.

¹⁸ "Meet Your New Assistant: Meta AI, Built With Llama 3." Meta, 18 Apr. 2024, about.fb.com/news/2024/04/meta-ai-assistant-built-with-llama-3.

¹⁹ Barca, Robert. "Údajná Nahrávka Telefonátu Predsedu PS a Novinárky Denníka N Vykazuje Podľa Expertov Početné Známky Manipulácie." AFP Fact Check, 28 Sept. 2023, fakty.afp.com/doc.afp.com.33WY9LF.

Sin embargo, la grabación se compartió durante un período crítico de 48 horas de silencio previo a las elecciones, durante el cual los candidatos tienen prohibido hacer campaña. Dado el momento, a los grupos de vigilancia les resultó difícil desacreditar adecuadamente la desinformación, ya que los ciudadanos ya la habían visto antes de acudir a las urnas. Si bien es difícil aislar el impacto que este evento tuvo en el resultado²⁰ de las elecciones, es importante que los actores de la democracia sigan rastreando y mapeando usos como este.



Caso de estudio

GLOBAL | EXPLOCIÓN DE IA PARA MANIPULACIÓN A ESCALA

En marzo de 2024, el Grupo Insikt de Recorded Future —una de las empresas de inteligencia de amenazas más grandes del mundo— identificó una red de influencia maligna que se aprovechaba de medios de comunicación no auténticos en Estados Unidos, el Reino Unido y Francia. Se sospecha que la red CopyCop está operada por Rusia, e Insikt Group informó que “la red utilizó ampliamente modelos de lenguaje extensos (LLM – Large Language Models por sus siglas en inglés) para plagiar, traducir y modificar contenido de fuentes de medios legítimos para adaptar mensajes políticos con sesgos específicos, que incluían contenido crítico de las políticas occidentales y de apoyo a las perspectivas rusas sobre cuestiones internacionales como el conflicto de Ucrania...”²¹ A medida que este contenido ganó fuerza, CopyCop comenzó a publicar también contenido producido por humanos junto con su contenido generado por IA como una estrategia para generar y distribuir contenido a gran escala para evocar emociones políticas específicas. The Economist informó que, en dos meses, CopyCop subió más de diecinueve mil publicaciones engañosas.²² La aparición de redes potenciadas por IAGen como esta crea más caos antes de elecciones críticas, lo que hace que la tarea de salvaguardar las elecciones sea aún más desafiante.

Exacerbación del discurso de odio y el acoso²³

La IAGen podría acelerar las tendencias existentes que han plagado los ecosistemas en línea, especialmente la propagación del discurso de odio, el acoso en línea y la violencia facilitada por la tecnología (incluidas las formas basadas en el género). Estos ataques se dirigen desproporcionadamente a las comunidades vulnerables y marginadas, como las mujeres y las personas no conformes con su género, los jóvenes, las personas LGBTQI+²⁴ las personas con discapacidad y otras minorías religiosas y raciales. La IAGen puede conducir a la creación de ataques sostenidos y automatizados, un mayor alcance del discurso de odio y el acoso más allá de las fronteras nacionales y la generación barata y fácil de contenido como deepfakes, imágenes sexuales no consentidas y suplantación de identidad, todo lo cual aumenta los riesgos y las amenazas, en línea y fuera de línea, para las comunidades enumeradas anteriormente.

²⁰ “Poll of Polls: Slovakian Polls, Trends and Election News for Slovakia.” POLITICO, www.politico.eu/europe-poll-of-polls/slovakia. Accessed 26 June 2024.

²¹ Insikt Group. “Russia-Linked CopyCop Uses LLMs to Weaponize Influence Content at Scale.” Recorded Future, 9 May 2024, www.recordedfuture.com/russia-linked-copycop-uses-llms-to-weaponize-influence-content-at-scale.

²² “A Russia-linked Network Uses AI to Rewrite Real News Stories.” The Economist, 10 May 2024, www.economist.com/science-and-technology/2024/05/10/a-russia-linked-network-uses-ai-to-rewrite-real-news-stories.

²³ The United Nations defines hate speech as any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor. See: United Nations. “What Is Hate Speech?” United Nations, www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech. Accessed 26 June 2024.

²⁴ Lesbian, gay, bisexual, transgender, queer, intersex, and more.

Como se hace referencia en el informe de la UNESCO sobre la violencia de género facilitada por la tecnología en la era de la IAGen,²⁵ ha habido evidencia de que los modelos de IA abiertos y cerrados se pueden utilizar para propagar los daños existentes de la violencia de género facilitada por la tecnología (VGFT), como la suplantación de identidad, la piratería informática y el acecho, y el acoso cibernético, entre otros. En conjunto, estos casos de uso perpetúan actitudes y prejuicios discriminatorios e imponen barreras de entrada al diálogo público y político para los grupos vulnerables y marginados, lo que afecta negativamente al pluralismo político.



Caso de estudio

GLOBAL | GENERANDO DEEPFAKES DIFAMATORIOS

DeepTrace, una empresa holandesa que rastrea contenido multimedia sintético, publicó un informe²⁶ en septiembre de 2019 que destaca el estado de los deepfakes creados por IAGen. De manera alarmante, su equipo estimó que aproximadamente el 96 por ciento de todos los videos deepfake encontrados en línea eran pornografía no consentida.

Su informe, que incluía datos de los cinco sitios web de pornografía deepfake más populares en el momento de su publicación, también señaló que todo el contenido presentaba a mujeres cuyas imágenes se usaban sin su permiso. Si bien estos videos explícitos a menudo presentaban a personas comunes, también se atacaba a figuras públicas en lo que generalmente era un intento de difamar su reputación profesional. Mujeres en una variedad de puestos, desde periodistas veteranas²⁷ hasta funcionarias gubernamentales,²⁸ han estado regularmente en el centro de las campañas habilitadas por IAGen creadas para expulsarlas de la vida pública. Las preocupaciones sobre cómo IAGen se usa y puede usarse para amplificar el acoso que enfrentan las mujeres, en particular, han seguido creciendo desde 2019, y este problema sigue siendo una preocupación para otras comunidades marginadas que enfrentan ataques despectivos de manera similar.

Desconfianza generalizada en el entorno de la información

Los usos maliciosos de la IAGen, como la producción barata de desinformación escrita o deepfakes, podrían reducir aún más la confianza pública en la fiabilidad de los entornos de información. Al mismo tiempo, es probable que el uso más benigno y de propósito general de la IAGen para crear grandes cantidades de contenido abarrote el espacio informativo, reduciendo la calidad de la información en línea. A medida que aumenta la conciencia pública de la IAGen, la duda y la desconfianza en la información en línea podrían manifestarse ampliamente. Una mayor comprensión y conciencia de los ciudadanos sobre los deepfakes puede dar lugar a un profundo escepticismo general en cuanto a la fiabilidad de la información sobre actores políticos, campañas, resultados electorales y más. Este escepticismo puede conducir a una disminución del compromiso democrático, ya que los ciudadanos se desilusionan cada vez más con los actores democráticos y, posteriormente, con los procesos democráticos. Según un informe de OpenAI publicado en enero de 2023,²⁹ los autores afirmaron que los LLM pueden crear propaganda más persuasiva e impactante que las técnicas que ya se utilizan y a una fracción del costo; los autores reconocieron que IAGen podría mejorar significativamente las operaciones de influencia que siembran aún más confusión y escepticismo.

²⁵ "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI." UNESCO, 2023, <https://unesdoc.unesco.org/ark:/48223/pf0000387483.locale=en>.

²⁶ Ajder, Henry, et al. "The State of Deepfakes: Landscape, Threats, and Impact." Deeptrace, Sept. 2019, regmedia.co.uk/2019/10/08/deepfake-report.pdf.

²⁷ Ayyub, Rana. "I Was the Victim of a Deepfake Porn Plot Intended to Silence Me." HuffPost UK, 21 Nov. 2018, www.huffingtonpost.co.uk/entry/deepfake-porn-uk_5bf2c126e4b0f32bd58ba316.

²⁸ Jankowicz, Nina. "I Shouldn't Have to Accept Being in Deepfake Porn." The Atlantic, 26 June 2023, www.theatlantic.com/ideas/archive/2023/06/deepfake-porn-ai-misinformation/674475.

²⁹ "Forecasting Potential Misuses of Language Models for Disinformation Campaigns and How to Reduce Risk." OpenAI, 11 Jan. 2023, openai.com/index/forecasting-misuse.

Los expertos han especulado que es probable que los gobiernos de la República Popular China y Rusia, así como actores nacionales, empresas e incluso personas influyentes, utilicen IAGen para crear contenido convincente que abruma las plataformas en línea en países específicos, desafiando los esfuerzos de moderación y reduciendo aún más la calidad de la información.



Caso de estudio

CHINA | EXPLOCIÓN DE HERRAMIENTAS DE IA DISPONIBLES PÚBLICAMENTE

En febrero de 2023, la empresa de análisis de redes sociales Graphika³⁰ publicó un informe³¹ sobre una operación de influencia a favor de la República Popular China que utilizaba presentadores de noticias generados por inteligencia artificial para difundir propaganda aprobada por el Estado. La campaña, a través de un medio ficticio llamado Wolf News, produjo un par de videos que circularon por las plataformas de redes sociales. El primer video incluía a un presentador que acusaba a Estados Unidos de no tomar medidas contra la violencia armada en el país, mientras que el segundo presentaba a un comentarista que enfatizaba la necesidad de cooperación entre Pekín y Washington. Cabe destacar que estos clips se generaron utilizando tecnología creada por Synthesia, una empresa con sede en el Reino Unido que crea personajes para que las empresas los utilicen como una forma de atraer al público objetivo de sus clientes. En su informe, Graphika señala que las herramientas disponibles públicamente como las proporcionadas por Synthesia han sido explotadas por actores malignos. Fuera de la República Popular China, los medios estatales de Venezuela también han utilizado la tecnología de Synthesia, sobre todo en 2023, cuando crearon un par de videos inventados en los que aparecían comentaristas de habla inglesa que promovían mensajes a favor del gobierno. Este incidente, que fue explorado en el informe Freedom on the Net 2023 de Freedom House,³² muestra cómo los actores autoritarios han aprovechado IAGen para crear propaganda.



Caso de estudio

GLOBAL | UN EJEMPLO DE GENERADORES DE DESINFORMACIÓN IMPULSADOS POR IA

En septiembre de 2023, un ingeniero que usaba el alias Nea Paw publicó un video en YouTube que detallaba su proyecto CounterCloud, un experimento en el que entrenaron a un LLM para producir contenido falso y engañoso.³³ Según The Debrief, el modelo fue diseñado para rastrear Internet en busca de contenido y, una vez que se seleccionaba un artículo, preparaba contra artículos atribuidos a periodistas falsos inventados por el sistema.³⁴ Estos medios sintéticos se publicaron luego en el sitio web basado en WordPress de CounterCloud, donde el modelo de IA creó personajes falsos para comentar sobre artículos específicos y simular la participación del usuario. Una vez creados los artículos, la IA se preparaba para compartir enlaces en X (anteriormente Twitter) y crear una serie de publicaciones, a menudo controvertidas, para despertar el interés. Si bien este experimento se llevó a cabo en un entorno cerrado, lo que significa que los artículos, personajes y publicaciones nunca estuvieron disponibles públicamente, el trabajo de Nea Paw llamó la atención sobre la facilidad con la que se puede utilizar IAGen para producir, difundir y atraer rápidamente el interés por desinformación convincente. Aunque la evidencia de estas tácticas aún es limitada, CounterCloud se presenta como un modelo barato y escalable que las autocracias pueden intentar emular.

³⁰ "Home." Graphika, graphika.com. Accessed 26 June 2024.

³¹ "Deepfake It Till You Make It." Graphika, Feb. 2023, public-assets.graphika.com/reports/graphika-report-deepfake-it-till-you-make-it.pdf.

³² "Freedom of the Net 2023: The Repressive Power of Artificial Intelligence." Freedom House, 2023, freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence#generative-ai-supercharges-disinformation.

³³ "CounterCloud - AI Powered Disinformation Experiment." YouTube, uploaded by neapaw713, 5 Sept. 2023, www.youtube.com/watch?v=cwGdkrc9i2Y.

³⁴ Banias, Mj. "Inside CounterCloud: A Fully Autonomous AI Disinformation System." The Debrief, 25 Nov. 2023, thedebrief.org/countercloud-ai-disinformation.

Ampliación de las brechas digitales

Según el informe Medición del desarrollo digital: hechos y cifras de la Unión Internacional de Telecomunicaciones (UIT), publicado en diciembre de 2023, 2.600 millones de personas, o un tercio de la población mundial, siguen sin conectividad.³⁵ En los países de bajos ingresos, menos personas están en línea y quienes usan muchos menos datos debido al acceso limitado e inestable y a las limitaciones de costos no pueden alcanzar el potencial total de la conectividad, como afirma la Internet Society.³⁶ Con la rápida evolución continua de los riesgos y capacidades de IAGen, quienes no tienen acceso a Internet, habilidades, recursos y herramientas para acceder, usar y aprovechar la IAGen de manera positiva se quedarán aún más atrás. Implementadas con intencionalidad, las soluciones de IAGen pueden mejorar el acceso y la inclusión y nivelar el campo de juego para la participación en el espacio digital y político, pero la opción predeterminada tiende a conducir a la exacerbación de las desigualdades existentes. Esta tendencia negativa puede llevar a que las campañas políticas, los actores de la sociedad civil y las personas en todo el mundo sigan enfrentando brechas de accesibilidad para usar modelos de IAGen. La brecha digital de género también puede amplificarse, abarcando desde el acceso limitado a las herramientas de IAGen para las mujeres hasta el hecho de que estas herramientas las afecten desproporcionadamente. Otros factores que resultan en la ampliación de las brechas digitales (que pueden afectar desproporcionadamente a algunos países del Sur Global, aunque muchos de estos desafíos también existen en los países del Norte Global) incluyen disparidades en la infraestructura de Internet, falta de herramientas políticas adecuadas para regular la IAGen y brechas en el uso de IAGen debido a la alfabetización limitada combinada con la dependencia de la IAGen en el formato de textos.

La escasa rendición de cuentas en el desarrollo y uso de IAGen genera una menor confianza

Los responsables de las políticas y los administradores electorales de todo el mundo tienen dificultades para seguir el ritmo del desarrollo y uso de la IAGen en los entornos electorales y de información. Este es especialmente el caso de algunos países del Sur Global que pueden enfrentar desafíos de infraestructura difíciles, recursos financieros limitados, una brecha digital más grande y una diversidad de barreras lingüísticas, dependencia extranjera de tecnologías externas o falta de marcos y regulaciones de rendición de cuentas existentes, entre otros factores.

Por ejemplo, como se ha visto tanto en el Norte como en el Sur Global, si un país ya tiene dificultades para exigir responsabilidades a los actores locales por la desinformación difundida en línea y fuera de ella, puede tener grandes dificultades para crear mecanismos de rendición de cuentas para las campañas políticas que hacen un uso indebido de la IAGen en las elecciones. Aunque se están haciendo algunos esfuerzos para regular la IA como se menciona en el manual, todavía no se han establecido las barreras básicas, lo que deja a los responsables de las políticas sin una hoja de ruta ni una guía para desarrollar regulaciones y una gobernanza efectivas de la IAGen. Aún más preocupante es el hecho de que, en ausencia de marcos que fomenten políticas basadas en derechos y que los respeten, la falta de marcos puede dar lugar a que los gobiernos aprueben leyes que hagan más daño que bien. Esta brecha y la falta de rendición de cuentas en torno al uso y el uso indebido de la IAGen crearán oportunidades para incentivar aún más a actores políticos malignos, incluidos gobiernos extranjeros y nacionales, a aprovechar el acceso barato y fácil a las herramientas de la IAGen para generar contenido manipulado que distorsione el ecosistema de información y sembramos la desconfianza.

“La hoja de ruta [para abordar las brechas digitales globales agravadas por la IAGen] requiere abordar tanto el impacto económico como el social de manera conjunta. Comienza desde el nivel de base y pasa por todos los ámbitos. Todos deben tener voz y voto en esta conversación.”

Lavina Ramkissoon (RAIN-África)

³⁵ “Measuring Digital Development – Facts and Figures 2023.” ITU, 2023, <https://www.itu.int/itu-d/reports/statistics/facts-figures-2023>. Accessed 19 July 2024.

³⁶ Aguiar, João. “One-Third of the Global Population Remains Offline.” Internet Society Pulse, 6 Dec. 2023, pulse.internetsociety.org/blog/one-third-of-the-global-population-remains-offline.

Mayores preocupaciones por la privacidad

Las empresas que construyen sistemas IAGen dependen de grandes cantidades de datos sobre y de individuos, incluyendo contenido protegido por derechos de autor y datos que contienen información personal. El rápido avance y uso de estos sistemas sin regulaciones de políticas orientadoras aumenta los riesgos de privacidad para los individuos y puede hacer que el público desconfíe de los usos de IAGen por parte del gobierno, la industria y la sociedad civil, que a menudo se superponen con la vigilancia impulsada por IA que algunos gobiernos utilizan. Según el Electronic Privacy Information Center, un centro de investigación de interés público con sede en Washington, DC, las comunidades marginadas se ven desproporcionadamente perjudicadas por las prácticas de recopilación de datos y los abusos de la privacidad tanto del gobierno como del sector privado, y se explotan a través de la vigilancia, el control y el sesgo algorítmico cuando se trata del uso de IA y aprendizaje automático.³⁷ La recopilación de datos sobre sus propios ciudadanos por parte de gobiernos con tendencias autoritarias también podría utilizarse con fines malignos para crear LLM autoritarios³⁸ y herramientas de IAGen para manipular a sus ciudadanos como método de control. La privacidad de los datos es un tema que rara vez se plantea ante el público y su importancia no se da a conocer, y hay muchos países donde la privacidad a menudo se considera un lujo en lugar de un derecho.

Concentración excesiva en soluciones técnicas para los riesgos de la IAGen

Las soluciones que se suelen proponer para abordar algunas de las amenazas descritas anteriormente son soluciones técnicas que se centran en adaptar o ajustar las herramientas IAGen. Por ejemplo, la idea, que suele debatirse, de confiar en la procedencia del contenido,³⁹ la inteligencia artificial, que ayuda a las personas a distinguir qué contenido es auténtico y cuál ha sido generado por IA, suele considerarse una solución técnica para un problema técnico que no tiene en cuenta la naturaleza sociopolítica de la desinformación. La tendencia a centrarse en soluciones técnicas para los problemas de la inteligencia artificial genómica (como el entrenamiento de modelos para reducir el sesgo,⁴⁰ la implementación de salvaguardas o el desarrollo de herramientas para detectar la desinformación de IAGen) resta importancia a las numerosas soluciones políticas, regulatorias y sociales que también están sobre la mesa.

Algunas de estas soluciones técnicas, como el establecimiento de sistemas de señalización de contenidos preocupantes, pueden además hacer que las personas afectadas tengan la responsabilidad de buscar recursos o incluso estos mecanismos pueden ser manipulados para usarlos contra las víctimas. Muchas soluciones no técnicas pueden, en última instancia, ser más eficaces, incluidas las campañas de educación pública, las redes de verificación de datos, el papel de los medios independientes en la creación de normas en torno a la información veraz y accesible y las regulaciones para las empresas de IAGen. Si uno se centrara únicamente en soluciones técnicas a los desafíos que genera la IAGen, podría correr el riesgo de no abordar las amenazas de manera integral.

“Esta estrategia [de invertir en iniciativas de alfabetización mediática] tendrá más poder de permanencia en el largo plazo, en lugar de aferrarse a herramientas de detección de inteligencia artificial que pueden no ser efectivas. Apoyar a las personas para que hagan su debida diligencia al lidiar con la desinformación puede ser la mejor manera de vacunar al público en general.”

**Representante de la industria,
en los Estados Unidos**

³⁷ “Privacy and Racial Justice.” Electronic Privacy Information Center, epic.org/issues/democracy-free-speech/privacy-and-racial-justice. Accessed 26 June 2024.

³⁸ The Decoder reported that in December 2023, the Chinese government released a dataset to train language models that reflect their political views and in line with its policies. See: Bastian, Matthias. “CCP Releases Politically Approved LLM Dataset With 50 Billion Tokens.” The Decoder, 29 Dec. 2023, the-decoder.com/ccp-releases-politically-approved-llm-dataset-with-50-billion-tokens.

³⁹ Content provenance refers to the basic, trustworthy facts about the origins of a piece of digital content (including images, videos, audio recordings, and documents). It may include information such as who created the content and how, when, and where it was created or edited. The content author always has control over whether provenance data is included as well as what data is included. Included information can be removed in later edits. Provenance also allows for anonymous content. See: “FAQ.” Coalition for Content Provenance and Authenticity, c2pa.org/faq. Accessed 26 June 2024.

⁴⁰ Training a large language model involves teaching it to understand and generate text by using a huge amount of written material, adjusting it to perform specific tasks or account for bias, and then testing and using it to do things like answer questions or write content. Model training to reduce bias can include using a diverse and representative dataset, removing sensitive variables, using bias mitigation techniques, and using human oversight.

OTHER OTROS FACTORES IMPORTANTES QUE CONDUCEN A UN AUMENTO DE LOS RIESGOS DE LA IA GENERATIVA PARA LA DEMOCRACIA

Para comprender mejor los riesgos descritos anteriormente, el IRI ha compilado una breve lista de factores que crean entornos propicios para que surjan esos riesgos. El objetivo es ayudar a los lectores a anticipar y minimizar los factores propicios y estar mejor preparados en entornos que probablemente permitan que se presenten algunos de los riesgos.

Falta de conciencia sobre los riesgos

En muchos países, los responsables de las políticas, los políticos, los funcionarios electorales y otros actores de la sociedad civil pueden desconocer las investigaciones existentes sobre la IAGen y sus posibles impactos. Esta falta de conciencia contribuye a una comprensión deficiente de los riesgos en general o del conocimiento de los riesgos y daños existentes. Los modelos IAGen también son intrínsecamente opacos y difíciles de entender, desde las fuentes de datos de los LLM utilizados en su desarrollo hasta la comprensión de cómo y por qué se genera un determinado resultado. Los desafíos para que el gobierno adquiera sistemas IAGen, la subcontratación del desarrollo y la integración de sistemas IAGen (por ejemplo, un gobierno que recurra a contratistas privados para la prestación de servicios públicos impulsados por IAGen) y el dejar las decisiones en manos de funcionarios públicos mal informados pueden ser factores que contribuyen o derivan de este mismo problema. A medida que los responsables de la toma de decisiones siguen luchando por desarrollar la alfabetización en IA en todo el mundo, especialmente en los países que luchan por superar las brechas de habilidades, la brecha en la comprensión se está ampliando, lo que magnifica una brecha digital existente. Las preocupaciones de los socios en los países del Sur Global en torno a las brechas digitales existentes han contribuido a las preocupaciones sobre el hecho de que los del Sur Global se queden atrás.

Opacidad en el diseño y el uso

Para los grupos de la sociedad civil, los gobiernos y los ciudadanos es difícil entender cómo funciona IAGen, ya que muchos modelos IAGen se consideran "cajas negras."⁴¹ Esto se ve agravado por el hecho de que algunas plataformas no articulan claramente sus políticas de contenido IAGen y sus acciones de aplicación, qué datos de entrenamiento se utilizan y cómo se diseñan o prueban los modelos. También es difícil para los actores de la sociedad civil y los ciudadanos buscar medidas de rendición de cuentas por el uso de IAGen por parte de los gobiernos y las campañas políticas cuando dicho uso suele ser secreto o subcontratado a empresas privadas. Lo que complica aún más esta situación es la ausencia de una definición clara de lo que significa la transparencia en IAGen y su implementación. Podría significar todo, desde revelar cómo se construye un modelo o publicar en qué datos de entrenamiento se basa, hasta dar al público más información sobre su diseño, desarrollo, implementación y mantenimiento. La falta de una comprensión establecida de lo que equivale a la transparencia deja a cada empresa privada en la obligación de definir la transparencia por sí misma.

Brechas de valor

Existen grandes brechas y variaciones entre los valores que guían a los desarrolladores de IA y los valores de los miles de millones de usuarios de todo el mundo, por lo que no es posible que las herramientas de IAGen globales reflejen una multitud de culturas y valores tan diversos. Muchas herramientas se desarrollan sin aportes diversos, lo que hace que sea difícil o poco práctico para los diferentes usuarios finales potenciales (que representan diversas culturas, idiomas e historias) utilizar ampliamente una herramienta de IAGen sin localizarla primero.

⁴¹ The lack of transparency and understanding of how AI systems make their decisions by the developers and users alike results in GenAI systems often being referred to colloquially as a "black box." As explained in a Vice article, "the people who develop AI are increasingly having problems explaining how it works and determining why it has the outputs it has." See: Xiang, Chloe. "Scientists Increasingly Can't Explain How AI Works." Vice, 1 Nov. 2022, <https://www.vice.com/en/article/y3pezm/scientists-increasingly-cant-explain-how-ai-works>.

Existe una amplia variación entre países como Brasil, India, México y Pakistán, por dar solo algunos ejemplos, que no se refleja en las herramientas IAGen "globales" no localizadas. Estas brechas afectan la forma en que se construye, utiliza y regula la IAGen de maneras importantes y relevantes para el contexto, y estas brechas permanecerán hasta que las herramientas sean localizadas por las diferentes personas que esperan usarlas.

"Los gobiernos tienen una responsabilidad adicional en materia de explicabilidad y transparencia, y deberían tener estas prioridades en mente antes de lanzar cualquier iniciativa."

Nicklas Badum (Democracy X)

Diversas políticas de las empresas y desafíos de moderación

La mayoría de las empresas de tecnología tienen políticas incipientes, en gran medida no probadas, sobre el uso de contenido la IAGen en contextos políticos o carecen de tales políticas en absoluto. Consulte el cuadro emergente a continuación para ver ejemplos de las herramientas IAGen más populares y sus políticas relevantes. Las políticas de las empresas son de amplio alcance y se actualizan de forma ad hoc, lo que dificulta el seguimiento y el monitoreo de las mismas por parte de los actores. A veces, tampoco está claro si las políticas de estas empresas se aplican a nivel mundial o solo en países seleccionados. Además, las políticas de consultas y resultados relacionados con las elecciones de las empresas (consulte el cuadro emergente a continuación), incluidas las fuentes que citan, se refieren principalmente a las elecciones del Norte Global e ignoran en gran medida el resto del mundo. Históricamente, las empresas también han invertido poco en la moderación de contenido en idiomas distintos del inglés. Esta falta de inversión, combinada con los despidos masivos del personal de la fundación y del personal, significa que la IAGen es especialmente impactante y riesgoso en un año en el que muchas elecciones se llevan a cabo en países donde el inglés no es el idioma predominante. Por último, es importante señalar que es extremadamente difícil monitorear la precisión de los resultados producidos por las herramientas de la IAGen dada la variedad e inconsistencia de las respuestas que las herramientas suelen proporcionar. Por ejemplo, un taller de AI Democracy Projects que investigó cómo los principales modelos de IA responden a las consultas que los votantes podrían hacer en enero de 2024 reveló que las respuestas de cinco modelos de IA líderes en ese momento (Claude de Anthropic, Gemini de Google, GPT-4 de OpenAI, Llama 2 de Meta y Mixtral de Mistral) a menudo eran inexactas, engañosas e incluso francamente dañinas.⁴² Esto probablemente será peor con entradas de texto que no estén en inglés o con la generación de salidas de texto que no estén en inglés porque la menor disponibilidad de datos para idiomas distintos del inglés puede resultar en una menor confiabilidad de los modelos de IA Gen que no estén en inglés.



Descripción general de las políticas de contenido actuales para las herramientas de IAGen de uso generalizado | Updated as of June 2024

A continuación, se presentan algunos de los generadores de contenido de IAGen que se utilizan más ampliamente a nivel mundial, junto con una lista de sus políticas y recursos confiables relacionados con las elecciones, el uso político y el etiquetado de contenido de IAGen y otras iniciativas clave. Esta lista no es exhaustiva y se centra en herramientas que tienen al menos algún tipo de políticas disponibles públicamente; la lista no captura todas las acciones que las plataformas están tomando para abordar este problema. Tenga en cuenta que existen muchas otras herramientas de IAGen creadas en otras regiones y países que también se utilizan ampliamente, pero que pueden ser accesibles solo para las personas que viven en esas localidades. Brookings publicó un artículo informativo de 2023 sobre las herramientas y los mercados de IAGen en el Sur Global.⁴³

⁴² Angwin, Julia. "Seeking Reliable Election Information? Don't Trust AI." Proof, 27 Feb. 2024, www.proofnews.org/seeking-election-information-dont-trust-ai/?ref=platformer.news.

⁴³ Okolo, Chinasa T. "AI In the Global South: Opportunities and Challenges Towards More Inclusive Governance." Brookings, 1 Nov. 2023, www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance.

Tenga en cuenta que estas políticas y recursos de confianza se actualizaron al momento de redactar este artículo, pero pueden cambiar a medida que la IAGen y los panoramas políticos y regulatorios sigan evolucionando. Recomendamos consultar periódicamente las páginas web de políticas de la herramienta. Al leer las políticas y los recursos que se enumeran a continuación, existe una gran variación en la forma en que se abordan (o no) los problemas políticamente sensibles y los procesos cívicos. Por último, tenga en cuenta que a medida que surgen regulaciones gubernamentales, pueden afectar las políticas de la empresa y requerir cambios, actualizaciones y, potencialmente, una mayor estandarización.



OpenAI ChatGPT: 1.800 millones de visitas mensuales⁴⁴

A principios de 2024, OpenAI dijo que ChatGPT remitiría a los usuarios a CanIVote.org para cualquier consulta relacionada con las elecciones estadounidenses. La política de la empresa también prohíbe la suplantación de identidad de candidatos, el uso de sus herramientas para hacer campaña, ejercer presión, desalentar la votación o tergiversar el proceso de votación.⁴⁵

OpenAI también anunció⁴⁶ que remitirá a los usuarios a la fuente oficial de información de votación del Parlamento Europeo para las elecciones al Parlamento Europeo de 2024.

OpenAI anunció que se unió al Comité Directivo de la Coalición para la Procedencia y Autenticidad del Contenido (C2PA)⁴⁷ en mayo de 2024 y comenzó a agregar metadatos de C2PA en ChatGPT para indicar las imágenes generadas por IA con un símbolo CR.



Google Gemini (anteriormente Bard): 414.4 millones de visitas mensuales

En diciembre de 2023, Google anunció que evitaría las consultas relacionadas con las elecciones estadounidenses y, en su lugar, remitiría a los usuarios a la Búsqueda de Google (esto se aplica a todas las consultas y resultados, no solo al texto).⁴⁸

En marzo 2024, Google anunció⁴⁹ que también restringirá los tipos de consultas relacionadas con las elecciones para las que Gemini devolverá respuestas sobre las elecciones indias y compartió que la compañía está trabajando con la Comisión Electoral de la India (ECI) para revelar información de calidad.⁵⁰

En mayo de 2024, Google anunció la ampliación de las capacidades SynthID⁵¹ de DeepMind para verificar no solo imágenes, sino también Gemini (texto) y Veo (video).

⁴⁴ Monthly visit information, referring to the average number of visits within a four-week period, were provided by Similarweb, a data aggregation company whose specialty is in website traffic and performance. Figures account for global users and reflect data from April 2024.

⁴⁵ "How OpenAI Is Approaching 2024 Worldwide Elections." OpenAI, 15 Jan. 2024, openai.com/index/how-openai-is-approaching-2024-worldwide-elections.

⁴⁶ "How OpenAI Is Approaching 2024 Worldwide Elections."

⁴⁷ C2PA is a widely used standard for digital content certification, developed and adopted by a wide range of actors including software companies, camera manufacturers, and online platforms. See: "FAQ." Coalition for Content Provenance and Authenticity, c2pa.org/faq. Accessed 26 June 2024.

⁴⁸ Jasper, Susan. "How We're Approaching the 2024 U.S. Elections." Google The Keyword, 19 Dec. 2023, blog.google/outreach-initiatives/civics/how-were-approaching-the-2024-us-elections.

⁴⁹ "Watermarking AI-generated Text and Video With SynthID." Google DeepMind, 14 May 2024, deepmind.google/discover/blog/watermarking-ai-generated-text-and-video-with-synthid.

⁵⁰ Google India Team. "Supporting the 2024 Indian General Election." Google India Blog, 12 Mar. 2024, blog.google/intl/en-in/company-news/outreach-initiatives/supporting-the-2024-indian-general-election.

⁵¹ "Watermarking AI-generated Text and Video With SynthID." Google DeepMind, 14 May 2024, deepmind.google/discover/blog/watermarking-ai-generated-text-and-video-with-synthid.



Perplexity AI: 70.54 millones de visitas mensuales

El motor de búsqueda basado en inteligencia artificial Perplexity AI anunció que sus algoritmos priorizan fuentes confiables y de buena reputación, como medios de noticias, y siempre proporciona enlaces para que los usuarios puedan verificar su resultado.⁵²

Los términos de servicio de Perplexity AI establecen que no está permitido usarlo para actividades ilegales, fraudulentas o dañinas, lo que incluye la manipulación política.⁵³ No se brindan detalles adicionales sobre las elecciones.



Anthropic Claude: 65.97 millones de visitas mensuales

La política de uso aceptable de Anthropic⁵⁴ prohíbe⁵⁵ el uso de sus herramientas en campañas políticas, cabildeo, incitación/facilitación de la interrupción de procesos electorales y cívicos, disuasión de la votación, etc. Los candidatos no podrán utilizar Claude para crear chatbots que puedan hacerse pasar por ellos o para campañas políticas específicas.

A Claude se le impide responder preguntas políticas de los usuarios estadounidenses, y en su lugar los dirige a TurboVote, una plataforma sin fines de lucro que alienta a los estadounidenses a votar al brindarles una guía de votación confiable y útil.



Microsoft Copilot (anteriormente Bing): 34.78 millones de visitas mensuales

En febrero de 2024, Microsoft utilizará capacidades de procedencia de medios basadas en métodos criptográficos para marcar y firmar contenido generado por IA con metadatos sobre su origen e historial, mediante el uso de las credenciales digitales de C2PA para marcar las imágenes generadas por IA con un símbolo especial.⁵⁶

En abril de 2024, Microsoft anunció la expansión de sus herramientas de integridad de contenido a los partidos políticos y campañas de la UE y a las organizaciones de noticias de todo el mundo.⁵⁷ Esto permite agregar credenciales de contenido⁵⁸ al contenido original y otorga a cualquier persona acceso a un sitio web⁵⁹ para verificar la existencia de credenciales de contenido en el contenido.

Microsoft también prohibirá a los usuarios que creen y distribuyan contenido engañoso y abusivo en sus servicios, como LinkedIn, la red de juegos de Microsoft y otros servicios relevantes.



Midjourney: 17.04 million monthly visits

Midjourney, un generador de imágenes con inteligencia artificial, no permite que su herramienta se utilice para generar imágenes para campañas políticas o para influir en el resultado de una elección.⁶⁰

⁵² Davis, Wes. "How AI Companies Are Reckoning With Elections." The Verge, 19 Mar. 2024, www.theverge.com/2024/3/19/24098381/ai-chatbots-election-misinformation-chatgpt-gemini-copilot-bing-claude?ref=platformer.news.

⁵³ "Terms of Service." Perplexity, 4 June 2024, www.perplexity.ai/hub/legal/terms-of-service.

⁵⁴ "Usage Policy." Anthropic, 6 June 2024, www.anthropic.com/legal/aup.

⁵⁵ "Preparing for Global Elections in 2024." Anthropic, 16 Feb. 2024, www.anthropic.com/news/preparing-for-global-elections-in-2024.

⁵⁶ Smith, Brad. "Combating Abusive AI-generated Content: A Comprehensive Approach." Microsoft On the Issues, 13 Feb. 2024, blogs.microsoft.com/on-the-issues/2024/02/13/generative-ai-content-abuse-online-safety.

⁵⁷ Hutson, Teresa. "Expanding Our Content Integrity Tools to Support Global Elections." Microsoft On the Issues, 22 Apr. 2024, blogs.microsoft.com/on-the-issues/2024/04/22/expanding-our-content-integrity-tools-to-support-global-elections.

⁵⁸ Content Credentials reveal information about the origin and history of content available for everyone to access. It lets users understand how much editing or manipulation the content went through. See "Home." Content Credentials, <https://contentcredentials.org/>. Accessed 26 June 2024.

⁵⁹ "Content Integrity." Microsoft, contentintegrity.microsoft.com. Accessed 26 June 2024.

⁶⁰ "Terms of Service." Midjourney, 7 Mar. 2024, docs.midjourney.com/docs/terms-of-service.

En marzo de 2024, los informes mostraron que Midjourney había comenzado a bloquear a los usuarios para que no crearan imágenes falsas de los dos principales candidatos de las elecciones estadounidenses de 2024. La empresa aún no ha delineado sus cambios de política en relación con las próximas elecciones.⁶¹



Meta AI: 6.53 millones de visitas mensuales

Meta anunció en noviembre de 2023 que exigiría a los anunciantes políticos que revelaran si utilizaban inteligencia artificial u otras técnicas digitales para crear anuncios publicados en sus plataformas.⁶²

Meta también ha prohibido el uso de sus herramientas de IAGen por parte de campañas políticas y anunciantes.⁶³ Meta anunció en abril de 2024 que comenzará a etiquetar una gama más amplia de contenido como "Hecho con IA" cuando detecte indicadores de imagen de IA estándar de la industria o informes de personas por sí mismas.⁶⁴



Mistral's Le Chat: 4.78 millones de visitas mensuales

Las condiciones de servicio de Mistral no contienen ninguna política relacionada con elecciones o manipulación política, pero señalan específicamente la prohibición de generar contenido que incite al odio, al acoso o a la violencia.⁶⁵



Stability AI (Stable Diffusion 3): 3.12 millones de visitas mensuales

Stability AI, que creó el generador de IA de texto a imagen Stable Diffusion, actualizó su política de uso aceptable en marzo de 2024 y prohíbe la creación o promoción de desinformación, campañas de desinformación a gran escala, anuncios políticos, propaganda y campañas de influencia.⁶⁶



También es importante señalar que en febrero de 2024, las principales empresas de tecnología se unieron para presentar un acuerdo tecnológico para combatir el uso engañoso de la IA en las elecciones de 2024.⁶⁷ La lista de empresas que se adhirieron a este acuerdo incluye a Amazon, Google, TikTok, entre otras.⁶⁸ El acuerdo es voluntario e incluye compromisos de las empresas de tecnología que van desde la creación de tecnología de marca de agua, la detección y etiquetado de contenido generado por IA y la evaluación de los modelos que sustentan el software de IA para identificar riesgos de abuso y aumentar la conciencia y la educación del público sobre la IA.⁶⁹ Sin embargo, el acuerdo carece de un mecanismo de rendición de cuentas y su enfoque parece ser de corto plazo, aplicándose solo a las elecciones que se celebrarán en 2024.

⁶¹ O'Brien, Matt. "AI Image-generator Midjourney Blocks Images of Biden, Trump." Associated Press, 13 Mar. 2024, apnews.com/article/midjourney-ai-imagegenerator-biden-trump-deepfakes-bc6c254ddb20e36c5e750b4570889ce1.

⁶² Clegg, Nick. "How Meta Is Planning for Elections in 2024." Meta, 28 Nov. 2023, about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024.

⁶³ Paul, Katie. "Meta Bars Political Advertisers From Using Generative AI Ads Tools." Reuters, 7 Nov. 2023, www.reuters.com/technology/meta-bar-political-advertisers-using-generative-ai-ads-tools-2023-11-06.

⁶⁴ Bickert, Monika. "Our Approach to Labeling AI-Generated Content and Manipulated Media." Meta, 5 Apr. 2024, about.fb.com/news/2024/04/metis-approach-to-labeling-ai-generated-content-and-manipulated-media.

⁶⁵ "Legal Terms and Conditions." Mistral AI, mistral.ai/terms. Accessed 26 June 2024.

⁶⁶ "Acceptable Use Policy." Stability AI, stability.ai/use-policy. Accessed 26 June 2024.

⁶⁷ "AI Elections Accord: A Tech Accord to Combat Deceptive Use of AI in 2024 Elections." AI Elections Accord, www.aielectionsaccord.com. Accessed 26 June 2024.

⁶⁸ "AI Elections Accord: A Tech Accord to Combat Deceptive Use of AI in 2024 Elections."

⁶⁹ Bond, Shannon. "Tech Giants Pledge Action Against Deceptive AI in Elections." NPR, 16 Feb. 2024, www.npr.org/2024/02/16/1232001889/ai-deepfakes-election-tech-accord.

OPORTUNIDADES DE LA IA GENERATIVA PARA LA DEMOCRACIA

A pesar de estas amenazas, la IAGen, cuando se utiliza de forma adecuada y responsable, puede proporcionar beneficios significativos a las democracias de todo el mundo. Aunque la IAGen recién ha alcanzado una escala que la ha convertido en una herramienta ampliamente accesible, los actores democráticos (en particular los encargados de la toma de decisiones y los representantes de la sociedad civil) están empezando a aprovecharla de formas novedosas e innovadoras. Las soluciones de IAGen pueden marcar la diferencia a la hora de promover la toma de decisiones inclusiva, predecir los impactos de las políticas e impulsar iniciativas de mejora de las competencias, por nombrar solo algunos ejemplos. En esta sección se analiza en profundidad cómo la IAGen puede ser una fuerza positiva en las sociedades democráticas, siempre que se tomen ciertas precauciones.

Fortalecimiento de las estrategias para proteger los entornos de información

Se ha hablado mucho sobre la capacidad de IAGen para crear desinformación a gran escala. Sin embargo, la IAGen también puede permitir a los actores democráticos abordar el contenido engañoso de manera más eficiente y eficaz. El examen de las formas en que IAGen puede usarse para fortalecer la integridad de la información se basa en la protección de los derechos individuales a la libertad de expresión. Un espacio saludable y sólido no significa silenciar la libertad de expresión, sino usar la tecnología para ayudar a promover las libertades y los principios democráticos. Las herramientas impulsadas por IAGen, como las creadas por la empresa estadounidense Reality Defender,⁷⁰ los sistemas de verificación de datos forman parte de este esfuerzo al verificar la autenticidad de los contenidos de audio, visuales o basados en imágenes, lo que permite a los responsables de la toma de decisiones y a los representantes de la sociedad civil identificar los contenidos inventados con mayor facilidad. En estos primeros casos, la IAGen simplifica el etiquetado, la eliminación y, en última instancia, la ralentización de la difusión de información errónea o desinformación. Además, la IAGen podría aprovecharse para automatizar este trabajo, lo que permitiría a los actores democráticos abordar rápidamente las falsedades compartidas por actores malignos. Los sistemas de verificación de datos, que pueden clasificar grandes cantidades de información, pueden detectar información errónea o desinformación de formas que los humanos no pueden reproducir. Los modelos impulsados por la IAGen pueden incluso proporcionar a los usuarios promedio contrapuntos precisos para ayudar a desacreditar la información falsa, una posible aplicación destacada por los investigadores del Laboratorio de Gobernanza e Inteligencia Artificial Responsable (GRAIL) de la Universidad de Purdue⁷¹ en un artículo de 2023 publicado por The Conversation.⁷² Los ejemplos que se enumeran aquí son solo el comienzo en lo que respecta a cómo la IAGen podría proteger la integridad de la información. A medida que el campo continúa evolucionando, estas herramientas impulsadas por la IAGen pueden desempeñar un papel más importante en la forma en que los actores democráticos abordan el complejo tema de la integridad de la información.

Impulsando una toma de decisiones ágil y eficiente

Durante muchos años, se han creado soluciones digitales para mejorar la colaboración entre los encargados de la toma de decisiones, sus socios en la sociedad civil y el público en general, con distintos grados de éxito. Ahora que la IAGen es más accesible, los encargados de la toma de decisiones y las partes interesadas de la sociedad civil están empezando a ver cómo las herramientas impulsadas por la IAGen podrían fortalecer los esfuerzos anteriores, aumentar su eficacia y ampliar su impacto. Las herramientas participativas de IAGen son especialmente prometedoras si se consideran las diversas formas en que podrían garantizar que se escuche a más partes interesadas. La IAGen puede utilizarse para simplificar documentos de políticas complejos en texto que los encargados de la toma de decisiones puedan interpretar y utilizar fácilmente.

⁷⁰ "From Dating to Democracy, AI-generated Media Creates Multifaceted Risks." Deloitte Insights, 19 Oct. 2023, www2.deloitte.com/xe/en/insights/focus/tech-trends/2024/reality-defender-helping-companies-with-ai-content-detection.html.

⁷¹ "Governance and Responsible AI Lab (GRAIL)." Purdue University, www.cla.purdue.edu/academic/polsci/research/labs/grail/index.html. Accessed 26 June 2024.

⁷² Schiff, Daniel S., and Kaylyn Jackson Schiff. "Generative AI Like ChatGPT Could Help Boost Democracy – if It Overcomes Key Hurdles." The Conversation, 7 Nov. 2023, theconversation.com/generative-ai-like-chatgpt-could-help-boost-democracy-if-it-overcomes-key-hurdles-212664.

Podría digerir grandes cantidades de datos sobre distintas demografías y comunidades, proporcionando nuevos conocimientos que pueden utilizarse para crear una formulación de políticas más receptiva. La IAGen también podría integrarse en chatbots que puedan responder de manera eficiente a las preguntas de una variedad de electores, incluidos aquellos con conocimientos digitales limitados o comprensión de la gobernanza. Incluso podría impulsar plataformas que sinteticen la retroalimentación de los electores de una manera que ayude a los funcionarios a tomar decisiones. La IAGen tiene inmensas posibilidades de hacer que el trabajo de gobernanza sea más abierto, inclusivo y receptivo a todos los interesados si se diseña y aplica democráticamente. Muchos en la comunidad democrática tienen un optimismo inicial sobre cómo la IAGen podría usarse para construir sistemas más participativos, receptivos y eficientes si se establecen las medidas de protección adecuadas.



Estudio de caso

TAIWÁN | FACILITAR LA ELABORACIÓN DE POLÍTICAS EN FORMA COLABORATIVA

En 2015, los responsables de la toma de decisiones en Taiwán utilizaron una solución de IAGen para solicitar las opiniones de los ciudadanos mientras elaboraban regulaciones para los servicios de transporte compartido. Polis,⁷³ se implementó una plataforma dinámica que recopila comentarios de los usuarios e identifica preocupaciones compartidas para aumentar la participación de los ciudadanos en el proceso de toma de decisiones. Las personas pudieron compartir sus ideas sobre qué políticas serían efectivas, que la plataforma analizó para identificar áreas de consenso. Después de completar este paso, Polis generó un conjunto de recomendaciones que guiaron el diseño de regulaciones, que se publicaron en un informe del Centro de Impacto Público de 2019.⁷⁴ El éxito de este experimento ha llevado a los líderes de Taiwán a seguir explorando el lugar de IAGen en los procesos de formulación de políticas. En 2023, el Ministerio de Asuntos Digitales del país se unió a la iniciativa de Asambleas de Alineación del Proyecto de Inteligencia Colectiva, cuyo objetivo es recopilar comentarios de los ciudadanos sobre los riesgos relacionados con la IA y compartir estas preocupaciones con los líderes de la industria. Al observar estos ejemplos, el uso de Polis por parte de Taiwán subraya cómo IAGen puede ayudar a crear políticas receptivas que aborden las necesidades más urgentes de los ciudadanos.⁷⁵

Mejora de la capacidad organizativa

Los responsables de la toma de decisiones en todo el mundo suelen carecer del tiempo, los recursos y el personal necesarios para responder adecuadamente a las necesidades de sus electores. Del mismo modo, las organizaciones de la sociedad civil se ven obstaculizadas por la falta de fiabilidad de la financiación, las limitaciones de capacidad y los retos logísticos a la hora de llevar a cabo tareas que requieren mucho tiempo. Los primeros hallazgos indican que la IAGen puede ser parte de la solución a estos problemas acuciantes, permitiendo a los actores democráticos hacer más con menos. Los responsables de la toma de decisiones y la sociedad civil podrían utilizar la IAGen para preparar materiales de comunicación, como puntos de debate o borradores de correos electrónicos, que sean adecuados tanto para el público interno como para el externo. Podría aprovecharse para resumir y sintetizar el contenido, extrayendo ideas que sean relevantes para el trabajo de los responsables de la toma de decisiones y la sociedad civil. En algunos casos, la IAGen podría utilizarse para mejorar la prestación de servicios a los electores que necesitan asistencia, como los que padecen inseguridad alimentaria, automatizando así procesos que pueden haber sido ineficientes.

⁷³ "Home." Polis, pol.is/home. Accessed 26 June 2024.

⁷⁴ "Building Consensus and Compromise on Uber in Taiwan." Centre for Public Impact, 18 Sept. 2019, www.centreforpublicimpact.org/case-study/building-consensus-compromise-uber-taiwan.

⁷⁵ Cushing Rodriguez, Sebastian. "Consensus Building in Taiwan, the Poster Child of Digital Democracy." Democracy Technologies, 4 Oct. 2023, democracy-technologies.org/participation/consensus-building-in-taiwan.

⁷⁶ Viamo. "Ask Viamo Anything." Viamo, viamo.io/ask-viamo-anything-ai. Accessed 26 June 2024.

Por ejemplo, Viamo, una plataforma basada en voz diseñada para áreas con baja conectividad, ha incorporado IAGen⁷⁶ para ayudar a responder las consultas de los usuarios sobre los servicios de atención médica. Si bien la eficacia de la IAGen depende de que los usuarios tengan las habilidades y los conocimientos necesarios para poder utilizarla, la IAGen promete mejorar el trabajo de los profesionales capacitados que navegan por una red de limitaciones y restricciones.

“La IA generativa ofrece una oportunidad para un gobierno conversacional. Los chatbots, por ejemplo, permiten responder las preguntas de los ciudadanos las 24 horas del día, los 7 días de la semana. En lugar de tener que esperar en largas colas o completar formularios en papel, pueden recibir ayuda inmediata.”

Beth Simone Noveck (centro Burnes para el cambio social en la Universidad Northeastern)



Caso de estudio

GLOBAL | AMPLIAR EL ACCESO A LOS SERVICIOS GUBERNAMENTALES

Muchos responsables de la toma de decisiones y grupos de la sociedad civil están considerando cómo la IAGen podría ampliar o simplificar la capacidad de los ciudadanos para acceder a los servicios. Un ejemplo temprano de esto es el uso de la IAGen para ayudar mejor a las sobrevivientes de la violencia doméstica a obtener el apoyo que necesitan. En 2022, la organización suiza sin fines de lucro Spring ACT lanzó su chatbot Sophia,⁷⁷ el cual es un producto que recoge evidencias presentadas por los usuarios, les brinda información actualizada sobre sus derechos y les ofrece un menú de opciones en función de sus respuestas. Un producto similar llamado SARA⁷⁸ fue creado en 2023 por Infosegura, una alianza estratégica entre el Programa de las Naciones Unidas para el Desarrollo (PNUD) y la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID). Al igual que Sophia, SARA es una herramienta que brinda a las mujeres y niñas que han sufrido violencia acceso a información sobre violencia de género, así como a los servicios de atención institucionales y de la sociedad civil disponibles. También ofrece asesoramiento legal, así como apoyo en la elaboración de planes de emergencia, entre otros servicios. Soluciones como estas sirven como ejemplos tempranos de cómo la sociedad civil puede usar la IAGen para ayudar a las personas a comprender qué servicios están disponibles para ellas en momentos de crisis.



Caso de estudio

ARGENTINA | SIMPLIFICAR LAS TAREAS DIARIAS

El proyecto Prometea,⁷⁹ que surgió de una colaboración entre la Facultad de Derecho de la Universidad de Buenos Aires y la Procuraduría General de la Nación, es un sistema impulsado por la IAGen diseñado para reducir la presión sobre los funcionarios públicos del sistema judicial municipal. En un artículo escrito por el líder del equipo universitario, esta novedosa herramienta logra esto de varias maneras.⁸⁰ En primer lugar, Prometea actúa como un asistente virtual para los funcionarios públicos, respondiendo a preguntas sobre determinadas leyes o estatutos. Sus desarrolladores señalan que su solución puede incluso responder a consultas específicas realizadas mediante comandos de voz.

⁷⁷ “Sophia Chatbot.” Spring ACT, springact.org/sophia-chatbot. Accessed 26 June 2024.

⁷⁸ “Home.” Sistema De Atención Y Resguardo De Información Automatizada (SARA), chatbotsara.org. Accessed 26 June 2024.

⁷⁹ Ast, Federico. “Prometea, Artificial Intelligence in the Judicial System of Argentina.” Medium, 14 Dec. 2021, medium.com/astec/prometea-artificial-intelligence-in-the-judicial-system-of-argentina-4dfbde079c40.

⁸⁰ Corvalan, Juan Gustavo. “Prometea: Artificial Intelligence to Transform Justice and Public Organizations.” International Journal of Digital and Data Law, vol. Vol 6, 2020, core.ac.uk/download/pdf/322501055.pdf.

Además, el sistema puede completar, e incluso generar, documentos legales sin la intervención de los usuarios. Al mismo tiempo, puede clasificar grandes cantidades de estos recursos en una fracción del tiempo que normalmente le llevaría a una persona hacerlo. Innovaciones como Prometea pueden resultar invaluable para las organizaciones que se enfrentan a limitaciones de capacidad. Con la participación humana, así como con una supervisión adecuada, los profesionales legales de todo el mundo pueden ver importantes ganancias de eficiencia.

Mejorar la transparencia institucional

Una crítica habitual a las soluciones de IAGen es que son “cajas negras”, lo que significa que es imposible determinar cómo y por qué las herramientas de la IAGen realizan análisis y producen resultados. Aunque la IAGen en sí es difícil de entender por completo, algunos en la comunidad democrática son tentativamente optimistas sobre su potencial para ser aprovechada como una herramienta que fortalezca los esfuerzos de transparencia. Hay algunas formas de hacerlo. Podría crear visualizaciones simples de datos complejos utilizados por los tomadores de decisiones de formas claras y comprensibles. La IAGen podría presentar esta información en un formato dinámico, lo que permitiría a los ciudadanos interpretar datos confusos con mayor facilidad y al mismo tiempo mostrarles cómo esta información informa (o no) la asignación de recursos y la toma de decisiones de los gobiernos, entre otras tareas clave.

También podría hacer más accesible la información almacenada en bases de datos expansivas utilizadas por los encargados de la toma de decisiones. Por ejemplo, DC Compass, un chatbot creado por funcionarios de Washington, DC, responde a las preguntas de los usuarios sobre cómo se gestionan sus datos.⁸¹ Si se aplica y diseña democráticamente, la IAGen tiene la promesa de arrojar luz sobre el funcionamiento interno de los sistemas de gobernanza y aumentar la transparencia y el acceso a información importante, incluso para los ciudadanos con bajo nivel de alfabetización digital y de datos en el Norte y el Sur Globales.



Caso de estudio

INDIA | AUMENTAR LA APERTURA GUBERNAMENTAL

La aplicación Digital Sansad⁸² lanzada en 2023 por los responsables de la toma de decisiones en la India, fue una forma de aumentar la transparencia en el parlamento. Una característica notable de esta solución impulsada por la IAGen es su capacidad de transcribir automáticamente los procedimientos gubernamentales a 22 idiomas mediante técnicas de reconocimiento de voz automatizado.⁸³ El equipo detrás de Digital Sansad permite a los usuarios abrir la aplicación y leer una transcripción de las deliberaciones en cuestión de minutos, lo que hace que la información y el proceso de gobierno sean mucho más accesibles para las personas de todo el país. Además, los desarrolladores de la herramienta planean integrar robots (bots) de chat y de voz que responderán las preguntas de los electores. Los usuarios podrían aprender fácilmente más sobre el trabajo de sus tomadores de decisiones en cualquier momento, incluso si su nivel de alfabetización digital puede ser relativamente bajo. Para los países donde las brechas digitales crean barreras a la accesibilidad y la participación de muchos ciudadanos, soluciones como Digital Sansad podrían marcar una diferencia significativa a la hora de aumentar la comprensión de los procesos gubernamentales por parte de los ciudadanos.

⁸¹ Edinger, Julia. “New Washington, D.C., Tool Uses Generative AI to Make Data Accessible.” GovTech, 26 Mar. 2024, www.govtech.com/artificial-intelligence/new-washington-d-c-tool-uses-generative-ai-to-make-data-accessible.

⁸² “Home.” Digital Sansad, sansad.in/phistory. Accessed 26 June 2024.

⁸³ Mathur, Swati. “‘Digital Sansad’ to Use AI to Transcribe House Proceedings Real-time.” The Times of India, 26 Apr. 2023, timesofindia.indiatimes.com/india/digital-sansad-to-use-ai-to-transcribe-house-proceedings-real-time/articleshow/99770031.cms.

Ampliación de los esfuerzos de educación cívica

Los responsables de la toma de decisiones y los grupos de la sociedad civil están experimentando con la forma en que la IAGen puede mantener al público informado sobre el funcionamiento de la gobernanza. Por ejemplo, existe la posibilidad de que los actores democráticos aprovechen las soluciones de IAGen para crear módulos de aprendizaje que se puedan adaptar fácilmente y que respondan a las aportaciones de los ciudadanos. El contenido de estos programas se puede modificar para cubrir las lagunas de conocimiento de los individuos, como lo que empresas como Teachflow⁸⁴ están considerando para mejorar la ciencia política y la educación cívica en los entornos escolares. Los módulos de educación cívica podrían informar a los usuarios sobre los procesos de gobernanza, así como sobre cómo podrían participar de manera significativa.

Además, la IAGen podría utilizarse para responder a preguntas de las personas relacionadas con políticas. Por ejemplo, los chatbots⁸⁵ impulsados por la IAGen se han utilizado para responder preguntas sobre determinados departamentos o normativas, proporcionar a los usuarios información actualizada y dirigirlos a los recursos pertinentes. A través de estas aplicaciones, la IAGen proporciona a las personas el conocimiento que necesitan para participar de manera significativa en los procesos de gobernanza. Si bien deben existir barandillas para evitar compartir desinformación, estos casos de uso muestran cómo la IAGen podría ayudar a garantizar que el público en general participe cívicamente y esté informado con precisión.



Caso de estudio

UNIÓN EUROPEA | MEJORAR LA DINÁMICA DE LOS CIUDADANOS Y TOMADORES DE DECISIONES

El Proyecto TITAN⁸⁶ que se puso en marcha en septiembre de 2022, es una iniciativa de múltiples partes interesadas puesta en marcha para ayudar a los ciudadanos cuando encuentran información en Internet que puede ser falsa o engañosa. TITAN es un chatbot impulsado por la IAGen diseñado para guiar a los usuarios en la evaluación de la precisión y fiabilidad de la información que se encuentra en las redes sociales. Uno de los objetivos de este proyecto es abordar los bajos niveles de confianza en las instituciones. El chatbot proporcionará a los usuarios información verificada sobre cómo trabajan para ellos los responsables de la toma de decisiones. El equipo de TITAN espera que los ciudadanos promedio puedan ganar confianza en sus responsables de la toma de decisiones, adquirir más conocimientos sobre las instituciones clave y sentirse más inclinados a participar en la vida cívica. Los resultados de este proyecto aún no están disponibles y todavía está por determinar si TITAN puede realmente fortalecer la confianza. Al igual que con la integración de cualquier herramienta de IAGen, puede haber riesgos o daños inesperados que aún no se han visto. La observación continua de este esfuerzo ayudará a demostrar las oportunidades (y los límites) del uso de estas herramientas para la educación cívica y los esfuerzos de creación de confianza.

Creación de ciberdefensas adaptables

Los defensores de la democracia, ya sean tomadores de decisiones o actores de la sociedad civil, han tenido que hacer frente durante mucho tiempo a ciberataques diseñados para descarrilar su trabajo. Si bien la IAGen ya ha comenzado a desempeñar un papel en las campañas lanzadas por actores autoritarios y de tendencia autoritaria, es importante señalar que la IAGen también podría ayudar a mejorar las contramedidas defensivas.

⁸⁴ "The Role of AI in School Political Science and Civics." Teachflow, 25 Jan. 2023, teachflow.ai/the-role-of-ai-in-school-political-science-and-civics.

⁸⁵ "KI-Bürgerassistenz Lumi Soll Nutzerinnen Und Nutzer Kennenlernen." Heidelberg, 24 Apr. 2022, www.heidelberg.de/HD/Presse/24_10_2022+ki-bu-ergerassistenz+lumi+soll+nutzerinnen+und+nutzer+kennenlernen.html.

⁸⁶ "Home." Titan, www.titanthinking.eu. Accessed 26 June 2024.

Las primeras investigaciones han demostrado que la IAGen podría utilizarse para rastrear amenazas potenciales, identificar vulnerabilidades explotables e incluso realizar análisis forenses, entre otras tareas críticas necesarias para una ciberdefensa sólida.⁸⁷ También podría ofrecer un apoyo fundamental a los actores con poca capacidad que se enfrentan a ataques habituales. Por ejemplo, podría impulsar un chatbot informativo que proporcione información sobre amenazas para ayudar a los actores democráticos a gestionar mejor sus recursos limitados al abordar amenazas cibernéticas urgentes. Fuera de las emergencias, también podría agilizar el trabajo necesario para defenderse de los ciberataques, reduciendo la carga que recae sobre los tomadores de decisiones y las organizaciones de la sociedad civil. Otro ejemplo es la IAGen, que podría proporcionar orientación sobre cómo evitar los ciberataques mediante simulaciones y ejercicios que imiten las tácticas y estrategias de los actores malignos para ayudar a preparar a los responsables de la toma de decisiones y a los líderes de la sociedad civil. Aunque se necesitan directrices sobre el uso de la IAGen para la ciberseguridad antes de que se aproveche para cualquiera de estos posibles casos de uso, su capacidad para abordar de forma proactiva las amenazas digitales es prometedora.



Caso de estudio

UCRANIA | SEGUIMIENTO DE POSIBLES ATAQUES CIBERNÉTICOS

Las empresas emergentes del sector tecnológico han estado prestando atención a cómo se podrían utilizar las soluciones IAGen para combatir las actividades en línea de actores autoritarios o con tendencias autoritarias. Un ejemplo de ello se puede ver en el trabajo de Mantis Analytics, un grupo con sede en Ucrania que ha creado una plataforma que monitorea la información compartida en línea, con el objetivo de exponer las falsedades que circulan en las redes sociales. Según la empresa, su solución puede identificar operaciones de influencia coordinadas que se encuentran en línea, al mismo tiempo que prepara contenido que corrige y aclara las falsedades. También puede detectar desinformación, ya que el sistema está entrenado para reconocer técnicas de manipulación comunes. La plataforma también ofrece análisis en tiempo real de las tendencias de la información en un panel interactivo, condensando el ruido en línea para que sea más digerible para los profesionales de la ciberseguridad. Mantis Analytics ha ayudado al Consejo Nacional de Seguridad y Defensa de Ucrania⁸⁸ en cuestiones de ciberdefensa y desde entonces ha recibido una financiación sustancial de Alchemist Accelerator, con sede en Estados Unidos.⁸⁹ Estos ejemplos indican cómo la IAGen, si se implementa de manera responsable, puede ayudar a las democracias a defenderse de ataques disruptivos en el ámbito digital.

Predicción de posibles resultados de políticas

Los responsables de la toma de decisiones, como los del estado de California, han considerado cómo se podrían implementar las soluciones IAGen para evaluar los posibles impactos de las soluciones de política antes de que se implementen por completo.⁹⁰ Por ejemplo, los responsables de la toma de decisiones y la sociedad civil pueden utilizar la IAGen para trazar diferentes escenarios si una política entrara en vigor. Al ejecutar modelos predictivos para pronosticar qué resultados pueden surgir, los responsables de la toma de decisiones podrían obtener información sobre los ajustes que podrían hacerse para mitigar las consecuencias negativas o no deseadas.⁹¹

⁸⁷ Ferrag, Mohamed Amine, et al. "Generative AI and Large Language Models for Cyber Security: All Insights You Need." arXiv, 21 May 2024, <https://doi.org/10.48550/arXiv.2405.12750>.

⁸⁸ Sobachynskyi, Rostyslav. "Ukrainian-made AI Platform Mantis Analytics Detects Russian Fakes in the News." AIN.Capital, 15 Sept. 2023, ain.capital/2023/09/15/ukrainian-mantis-analytics-detects-russian-fakes.

⁸⁹ Laoun, Joy. "Ukrainian Mantis Analytics Secures \$30k Investment From the Alchemist Accelerator." Vestbee, 11 Apr. 2024, www.vestbee.com/blog/articles/ukrainian-mantis-analytics-secures-a-30k-investment-from-the-alchemist-accelerator.

⁹⁰ State of California. State of California Report: Benefits and Risks of Generative Artificial Intelligence. Nov. 2023, www.govops.ca.gov/wp-content/uploads/sites/11/2023/11/GenAI-EO-1-Report_FINAL.pdf.

⁹¹ Digital Twins Digest. "Simulations 2.0: The Role of Generative AI in Creating Accurate and Reliable Models." Medium, 26 June 2023, Archived Version, <https://web.archive.org/web/20240604131848/https://medium.com/@thedigitalwindigest/simulations-2-0-the-role-of-generative-ai-in-creating-accurate-and-reliable-models-b6e45d91d0ba>.

Los responsables de la toma de decisiones también podrían aprovechar la IAGen para detectar tendencias y patrones emergentes dentro de conjuntos de datos más amplios que pueden determinar los impactos generales de una intervención. La sociedad civil podría aprovechar la IAGen para detectar posibles deficiencias en los análisis de políticas de los responsables de la toma de decisiones. Estas aplicaciones demuestran cómo la IAGen podría ser clave para abordar de manera proactiva los problemas de políticas antes de que surjan. No solo daría a los responsables de la toma de decisiones la oportunidad de corregir el rumbo antes de que sus electores se vean afectados, sino que también brindaría una oportunidad única para que la sociedad civil ofrezca una crítica más matizada de las intervenciones políticas.



Caso de estudio

NIGERIA | SATISFACER LAS NECESIDADES DE LA COMUNIDAD

El análisis de sentimientos, que se refiere a un proceso en el que se evalúa sistemáticamente el contenido basado en texto de muchas fuentes diferentes para determinar el tono general, ofrece una perspectiva prometedora para los tomadores de decisiones interesados en utilizar la IAGen para comprender mejor las necesidades de sus electores. Sin embargo, el desarrollo de herramientas que permitan el análisis de sentimientos en el Sur Global a menudo se ve obstaculizado por la falta de recursos lingüísticos para los desarrolladores de la IAGen que se encuentran en regiones de habla no inglesa. Para abordar esta brecha, Masakhane, una comunidad de base para expertos en procesamiento del lenguaje natural (PLN) en toda el África subsahariana, ha buscado crear un corpus y un léxico lingüísticos para hausa, yoruba e igbo,⁹² tres de los idiomas más hablados en Nigeria. Con este proyecto, Masakhane busca generar los activos necesarios para permitir la creación de herramientas para realizar análisis de sentimientos en sus propias comunidades. La aplicación de la IAGen para el análisis de sentimientos podría dar como resultado la reformulación de políticas basadas en las preocupaciones de los electores antes de que se implementen. Aunque el desarrollo de activos lingüísticos sigue siendo un desafío, Masakhane ofrece un sólido ejemplo de cómo los esfuerzos impulsados por la comunidad pueden funcionar para hacer posibles aplicaciones innovadoras de IAGen para el bien.

⁹² "NaijaNLP: Sentiment Lexicon and Hate Speech." Masakhane, www.masakhane.io/ongoing-projects/naijanlp-sentiment-lexicon-hate-speech. Accessed 26 June 2024.

RECOMENDACIONES

Las organizaciones de la sociedad civil (OSC), los gobiernos y las empresas del sector privado (desde las organizaciones no gubernamentales (ONG) de la democracia y los organismos de control de la rendición de cuentas hasta las administraciones electorales y las plataformas de redes sociales) tienen muchas oportunidades de maximizar el potencial de la IAGen para beneficiar a la democracia y, al mismo tiempo, mitigar sus riesgos. Esta subsección presenta las múltiples recomendaciones concretas que surgieron del grupo de trabajo del IRI para alinear el desarrollo, la implementación y el uso de la IAGen con los valores democráticos. Las recomendaciones se dividen en tres sectores clave: sociedad civil, gobierno e industria. Cabe señalar que, si bien las recomendaciones están organizadas por sector, muchas de ellas se refieren a la colaboración intersectorial que se necesita para minimizar eficazmente los daños y maximizar las oportunidades que presenta la IAGen. En estos esfuerzos se deben incluir voces y perspectivas diversas de manera predeterminada.

En todas las áreas, los miembros del grupo de trabajo fueron claros en que las OSC, los gobiernos y las empresas deben asegurarse de realizar consultas públicas y de divulgación inclusivas para buscar, escuchar y enfatizar las perspectivas de las personas, las comunidades y las regiones que enfrentan la marginación y las desigualdades en la sociedad en general. Más importante aún, es fundamental que estas perspectivas se incorporen de manera efectiva en las acciones que surjan como resultado de estos aportes. Esto incluye considerar cómo factores como la raza, el género, la edad, el nivel socioeconómico, la capacidad, la primera lengua, el país y la región global (por ejemplo, los países del Sur Global y la variación dentro del Sur Global) afectan las oportunidades y los riesgos de la IAGen para esas personas.

RECOMENDACIONES PARA LA SOCIEDAD CIVIL

Las ONG que trabajan en pro de la democracia, las organizaciones de verificación de datos, los organismos de control de la rendición de cuentas (en los medios de comunicación, la política, la tecnología y otras áreas) y otras OSC pueden adoptar las siguientes medidas para contribuir a garantizar que la IAGen se utilice de forma democrática dentro de sus propias organizaciones y por parte de los gobiernos, las empresas y los ciudadanos de sus propios países:



Antes de tomar cualquier otra medida, mejore sus habilidades y las de su equipo en los conceptos básicos de la IA generativa. Es importante entender qué es exactamente la IAGen y los riesgos y amenazas que puede presentar para ver dónde, si es que lo hace, puede ser útil para el trabajo de uno. Recuerde que aprovechar la IAGen puede no ser aplicable a todas las organizaciones, y que muchas soluciones a los desafíos de gobernanza o sociales son analógicas en lugar de digitales. Consulte la sección de introducción que aparece al principio del informe, que enumera recursos educativos y documentos explicativos para ayudar a desarrollar una comprensión básica de las herramientas mejoradas con la IAGen. Después de aprender los conceptos básicos, considere **cómo integrar la IAGen en el trabajo de su organización o educar y promover los riesgos y oportunidades que presenta la IAGen.** Las siguientes recomendaciones abordan y están segmentadas en función de esas dos necesidades.

Integración y aprovechamiento de la IAGen en su organización

Consulte los marcos de uso de la IAGen basados en derechos para desarrollar políticas internas de uso de la IAGen.

Apoyarse en marcos como las directrices de Estados Unidos⁹³ y Nueva Zelanda⁹⁴ para usos de agencias gubernamentales, las directrices del Foro del Área de Investigación Europea⁹⁵, y los principios éticos de la IA de las Naciones Unidas⁹⁶, puede informar la creación de políticas internas de uso de la IAGen que deberían dejar en claro al personal y socios de las OSC cómo deben o no deben usar la IAGen en sus trabajos diarios.

Por ejemplo, una política podría permitir que el personal utilice la IAGen para la traducción de idiomas, siempre que haya una verificación humana posterior, y también prohibir que el personal y los socios introduzcan datos personales en los modelos de IAGen. Los grupos de la sociedad civil podrían publicar estas políticas en sus sitios web para lograr transparencia y también como referencia para otros. El objetivo de adoptar estos marcos debería ser identificar y permitir usos positivos y efectivos de la IAGen internamente, capacitar y educar al personal, contribuir a los estándares de uso de la IAGen dentro de la sociedad civil y aumentar la rendición de cuentas por las prácticas de la sociedad civil. Por ejemplo, el Responsible AI Institute, una organización sin fines de lucro que trabaja con actores clave que utilizan herramientas de IA, incluida la sociedad civil, ha desarrollado un conjunto de mejores prácticas⁹⁷ para ayudar a guiar a las organizaciones en su intento de integrar la IAGen de manera efectiva y apropiada.

Únase a las coaliciones tecnológicas globales (si aún no participa) y abogue en esos foros por un mayor enfoque en la IAGen y las regulaciones y gobernanza de la IA.

Considere la posibilidad de unirse a coaliciones como la organización sin fines de lucro Partnership on AI, la Tech Accountability Coalition del Aspen Institute y la Global Partnership on Artificial Intelligence (GPAI), apoyada por la OCDE. Si las OSC encuentran el uso de la IAGen en su trabajo diario, deberían compartir esos ejemplos y estudios de caso con otras OSC y, si es posible, publicarlos de alguna forma (por ejemplo, escribir blogs de estudios de caso, mencionar los casos de uso en seminarios web y en actividades de promoción pública) para ayudar a educar a los responsables de las políticas que trabajan en IAGen y democracia. Y cuando las OSC se presenten ante los responsables de las políticas para hablar sobre IA, la IAGen y temas relacionados, deberían aprovechar estas coaliciones tecnológicas globales para obtener nuevas investigaciones, artículos de noticias, mejores prácticas y otros materiales. Por ejemplo, la sociedad civil interesada en la transparencia y la auditoría de la IAGen podría beneficiarse de consultar el "Índice de Transparencia del Modelo de Fundación"⁹⁸ de investigadores de la Universidad de Stanford, el MIT y la Universidad de Princeton. También es de interés el informe de estudio de caso de Open Loop⁹⁹ sobre los esfuerzos de transparencia de la IA entre las empresas en México. Es fundamental que las organizaciones no gubernamentales internacionales y los propios organizadores de estas coaliciones reconozcan las barreras a la participación e identifiquen formas de incorporar las voces de la sociedad civil. Los grupos de la sociedad civil de alto perfil deberían abogar por la reducción intencional de las barreras a la participación, como el idioma, el estatus y las expectativas de reconocimiento de marca.

Desarrolle un plan de respuesta a incidentes que incorpore el uso de la IAGen. Un plan de respuesta a incidentes es, por lo general, un documento escrito que incluye una guía sobre cómo su organización puede prepararse antes y cómo actuar durante y después de un incidente de seguridad.

⁹³ "Responsible Use of Generative Artificial Intelligence for the Federal Workforce." U.S. Office of Personnel Management, www.opm.gov/data/resources/ai-guidance. Accessed 26 June 2024.

⁹⁴ "System Leaders' Guidance for Use of GenAI Across the New Zealand Public Service." New Zealand Government, www.digital.govt.nz/assets/Standards-guidance/Technology-and-architecture/Generative-AI/Joint-System-Leads-tactical-guidance-on-public-service-use-of-GenAI-summary-September-2023.pdf. Accessed 26 June 2024.

⁹⁵ Directorate-General for Research and Innovation. "Guidelines on the Responsible Use of Generative AI in Research." European Union, 20 Mar. 2024, research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/guidelines-responsible-use-generative-ai-research-developed-european-research-area-forum-2024-03-20_en.

⁹⁶ "Principles for the Ethical Use of Artificial Intelligence in the United Nations System." UN System Chief Executives Board for Coordination, 27 Oct. 2022, unsceb.org/sites/default/files/2023-03/CEB_2022_2_Add.1%20%28AI%20ethics%20principles%29.pdf.

⁹⁷ Responsible Artificial Intelligence Institute. "Best Practices in Generative AI Guide." Responsible AI, 2024, www.responsible.ai/best-practices-in-generative-ai-guide.

⁹⁸ Bommasani, Rishi, et al. "The Foundation Model Transparency Index." arXiv, 19 Oct. 2023. <https://doi.org/10.48550/arXiv.2310.12941>.

⁹⁹ Del Pozo, Claudia May, et al. "AI Transparency and Explainability (Mexico): Public Policy Prototype on the Transparency of Artificial Intelligence Systems." Open Loop, Aug. 2023, openloop.org/programs/ai-transparency-explainability-mexico.

Al hacerlo, se asegura de que su organización esté preparada en caso de incidentes cibernéticos o riesgos vinculados a la IAGen, como un producto IAGen que filtre datos confidenciales de la organización¹⁰⁰ de modo que el personal tenga una hoja de ruta sobre cómo responder. Debido a que la IAGen no existe en el vacío, las organizaciones que ya tienen un plan de respuesta a incidentes de ciberseguridad o medidas de ciberseguridad en general (por ejemplo, filtrado de phishing, firewalls) pueden evaluar cómo las posibles fugas de datos a través de productos IAGen, los avisos maliciosos¹⁰¹ que eluden las reglas del modelo IAGen y otros riesgos encajan en su plan de respuesta a incidentes. Para las organizaciones que no tienen un plan de respuesta a incidentes cibernéticos (o que tienen uno menos desarrollado), el Instituto Nacional de Estándares y Tecnología de EE. UU. proporciona un punto de partida con los tres pasos de su Ciclo de respuesta a incidentes:¹⁰² detectar, responder y recuperarse. La Agencia de Infraestructura y Seguridad Cibernética de Estados Unidos también explica claramente cómo las organizaciones deben crear un documento escrito de plan de respuesta a incidentes¹⁰³ e implementarlo. Y la Asociación Internacional de Profesionales de la Privacidad también publicó un artículo¹⁰⁴ sobre planes de respuesta a incidentes de IA con pasos útiles a seguir cuando los sistemas de IA han causado daños.

Sensibilización, promoción y educación sobre la IAGen

Denunciar los usos de la IAGen por parte de los gobiernos y el sector privado que violen, socaven o puedan comprometer los derechos humanos¹⁰⁵ como la privacidad y la libertad de expresión. Este podría ser un esfuerzo continuo de una amplia gama de OSC, incluidas aquellas con una inversión específica o un enfoque en la programación y la promoción de la gobernanza de la IA. Esto también es relevante para quienes invierten en la defensa de los grupos vulnerables, y en otros casos se realizará de manera ad hoc a medida que las organizaciones se topan con usos opacos, riesgosos o dañinos de la IAGen en su trabajo diario. Por ejemplo, Access Now, una organización sin fines de lucro dedicada a promover los derechos digitales, tiene una amplia experiencia en la defensa contra el uso de herramientas de IA que socavan las libertades individuales¹⁰⁶ en los últimos años. Freedom House también ha llamado la atención sobre cómo la IA puede limitar las libertades en Internet e impactar en las democracias en general.¹⁰⁷ Busque directrices o acuerdos existentes que ayuden a informar o respaldar su labor de promoción, incluso considerando declaraciones regionales que puedan estar más cerca de las prioridades de su organización, como la Declaración de Montevideo sobre Inteligencia Artificial y su Impacto en América Latina, entre otras.¹⁰⁸

“Estas tecnologías [impulsadas por IAGen] están beneficiando a las personas a nivel local. La creación de códigos de conducta para las campañas, así como la colaboración con organizaciones de vigilancia, son ejemplos de algunas soluciones que podríamos considerar para abordar este problema [del uso indebido de IAGen]”.

Julia Brothers (Instituto Nacional Democrático)

¹⁰⁰ Muncaster, Phil. “Fifth of CISOs Admit Staff Leaked Data via GenAI.” Infosecurity Magazine, 24 Apr. 2024, www.infosecurity-magazine.com/news/fifth-cisos-staff-leaked-data-genai.

¹⁰¹ Zhu, Banghua, et al. “Generative AI Security: Challenges and Countermeasures.” arXiv, 20 Feb. 2024, arxiv.org/abs/2402.12617.

¹⁰² “Incident Response.” NIST Computer Security Resource Center, csrc.nist.gov/Projects/incident-response. Accessed 26 June 2024.

¹⁰³ “Incident Response Plan (IRP) Basics.” Cybersecurity and Infrastructure Security Agency, 31 Jan. 2024, www.cisa.gov/resources-tools/resources/incident-response-plan-irp-basics.

¹⁰⁴ Leong, Brenda, and Daniel Atherton. “AI Incident Response Plans: Not Just for Security Anymore.” IAPP, 20 Sept. 2023, iapp.org/news/a/ai-incident-response-plans-not-just-for-security-anymore.

¹⁰⁵ United Nations General Assembly. The Universal Declaration of Human Rights (UDHR). New York: United Nations General Assembly, 1948, www.un.org/en/about-us/universal-declaration-of-human-rights.

¹⁰⁶ “Artificial Intelligence and Human Rights.” Access Now, <https://www.accessnow.org/issue/artificial-intelligence>. Accessed 19 July 2024.

¹⁰⁷ “Freedom of the Net 2023: The Repressive Power of Artificial Intelligence.”

¹⁰⁸ Various Authors. Montevideo Declaration on Artificial Intelligence and its Impact in Latin America. KHIPU, Mar. 2023, <https://zenodo.org/records/8208793>.

Apoyar los esfuerzos de educación social, como las campañas de educación cívica, en múltiples frentes, incluida la capacitación del personal interno. Identifique otros grupos de la sociedad civil que trabajen en programas de capacitación¹⁰⁹ para compartir estudios de casos, experiencias de primera mano con IAGen y otros recursos e información que puedan servir de base para sus esfuerzos de formación y educación con poblaciones de todo el mundo. Desarrolle materiales educativos basados en las propias experiencias de sus organizaciones con IAGen. Por ejemplo, una ONG de derechos humanos podría escribir blogs explicativos sobre IAGen y sus impactos en los derechos humanos (como lo hizo Access Now¹¹⁰), o una organización sin fines de lucro dedicada a la rendición de cuentas en materia de tecnología podría hablar sobre el contenido creado con IA que ve en las redes sociales (como lo hizo el Software Freedom Law Center en India).¹¹¹

La Universidad de Wollongong en Australia ha desarrollado recursos que resumen y explican¹¹² el impacto de IAGen y la IA en la educación, y el grupo de expertos con sede en Berlín Democracy Reporting International ha creado y publicado una guía sobre cómo reconocer mejor¹¹³ el contenido generado por IA en Internet. Basándose en estos recursos, las OSC también deberían utilizar sus redes y su posicionamiento fuera de la industria y el gobierno para amplificar las perspectivas y necesidades de diferentes edades, habilidades y grupos geográficos durante la capacitación, como lo hizo la UNESCO¹¹⁴ con su informe sobre los impactos de la IA en las comunidades indígenas de América Latina y el Caribe.

Además de invertir en iniciativas de formación externa, las OSC deberían ofrecer formación periódica a su personal sobre las promesas y los inconvenientes del uso de la IAGen e invertir en la formación de otras OSC (por ejemplo, mediante sesiones de formación conjuntas) sobre las formas de utilizar mejor la tecnología y hablar con los grupos de interés sobre IAGen. Incluso es posible que las OSC consideren la posibilidad de perfeccionar los modelos de lenguaje extensos (LLM) para satisfacer las necesidades específicas de la sociedad civil, como se ha hecho¹¹⁵ con el LLM Bloom de Hugging Face.

Busque leyes y regulaciones en los países en los que opera su organización para entender a qué datos relacionados con IAGen tiene acceso. Algunas regulaciones emergentes brindan cada vez más a la sociedad civil acceso a los datos de las empresas de tecnología para fines de investigación. Por ejemplo, la Ley de Servicios Digitales recientemente implementada por la UE requiere legalmente que las empresas¹¹⁶ consideradas “plataformas en línea muy grandes” y “motores de búsqueda muy grandes” brinden a investigadores seleccionados acceso a datos que contribuyan a detectar, identificar y comprender los riesgos sistémicos en los países de la UE. Las OSC que buscan comprender mejor cómo la IAGen impacta la desinformación, el discurso de odio, la prestación de servicios públicos y más podrían usar estos derechos de acceso a la investigación para estudiar datos no públicos y publicar perspectivas únicas sobre el espacio problemático.

¹⁰⁹ “InnovateUS Workshop Series Artificial Intelligence for the Public Sector.” InnovateUS, <https://innovate-us.org/workshop-series/artificial-intelligence-for-the-public-sector>. Accessed 26 June 2024.

¹¹⁰ Leufer, Daniel, and Méabh Maguire. “What You Need to Know About Generative AI and Human Rights.” Access Now, 24 May 2023, www.accessnow.org/what-you-need-to-know-about-generative-ai-and-human-rights.

¹¹¹ “Joint Letters to ECI and Platform Companies on Generative AI and Manipulated Media Content.” Software Freedom Law Center (India), 29 Mar. 2024, sfic.in/joint-letters-to-eci-and-platform-companies-on-generative-ai-and-manipulated-media-content.

¹¹² “AI in Education: Publicly Available genAI Tools.” University of Wollongong, 19 Mar. 2024, itc.uow.edu.au/hub/page/ai-resources-public-generative-ai-tools.

¹¹³ “Synthetic Media Exposed: DRI’s Comprehensive Guide to AI Disinformation Detection.” Democracy Reporting International, 10 Apr. 2023, democracy-reporting.org/en/office/global/publications/synthetic-media-exposed-dri-comprehensive-guide-to-ai-disinformation-detection.

¹¹⁴ “Inteligencia Artificial Centrada En Los Pueblos Indígenas: Perspectivas Desde América Latina Y El Caribe.” UNESCO, 2023, unesdoc.unesco.org/ark:/48223/pf0000387814.

¹¹⁵ Shixiang Research. “Deep Dive on Hugging Face.” Medium, 14 Nov. 2022, <https://medium.com/@shixiangtech2019/deep-dive-on-hugging-face-87cfd122c1a4>. Archived Version, <https://web.archive.org/web/20221114151347/https://medium.com/@shixiangtech2019/deep-dive-on-hugging-face-87cfd122c1a4>.

¹¹⁶ “FAQs: DSA Data Access for Researchers.” European Centre for Algorithmic Transparency, 13 Dec. 2023, algorithmic-transparency.ec.europa.eu/news/faqs-dsa-data-access-researchers-2023-12-13_en.

RECOMENDACIONES PARA LOS GOBIERNOS

Los gobiernos pueden adoptar las siguientes medidas para contribuir a garantizar que la IAGen se utilice de forma democrática tanto a nivel interno como al servicio de los valores públicos y democráticos. Estas recomendaciones podrían ser implementadas por diferentes elementos de los gobiernos, como, por ejemplo, las agencias de prestaciones sociales que atienden al público y que podrían utilizar un chatbot para mejorar la prestación de servicios a los ciudadanos, y los legisladores que podrían estar redactando leyes y dando forma a las regulaciones en torno a la IAGen:



Antes de tomar cualquier otra medida, es necesario capacitar a los funcionarios y al personal del gobierno sobre los conceptos básicos de la IA generativa. Es especialmente importante que el gobierno se tome el tiempo necesario para asegurarse de que exista una comprensión sólida y básica de qué es la IAGen y qué riesgos y amenazas puede presentar, a fin de entender en qué casos, si es que puede ser útil, para un proceso o equipo específico. Recuerde que aprovechar la IAGen puede no ser aplicable a diversos procesos gubernamentales, y que muchas soluciones a los desafíos de gobernanza o sociales son analógicas en lugar de digitales. Consulte la sección de introducción que aparece anteriormente en el informe, en la que se enumeran diferentes recursos educativos y documentos explicativos para ayudar a desarrollar conocimientos básicos sobre esta herramienta.

Después de aprender los conceptos básicos, los funcionarios del gobierno pueden identificar formas de integrar la IAGen en su propio trabajo, revisar o desarrollar políticas y regulaciones, o lanzar iniciativas de educación y concientización para garantizar que los ciudadanos estén informados sobre los riesgos y las oportunidades que presenta la IAGen. Las siguientes recomendaciones abordan estas diferentes necesidades.

Invertir en una descripción y notificación clara, comprensible y completa de cómo y cuándo se utiliza la IAGen en los servicios o productos del gobierno. Esto puede hacerse mediante la adición de etiquetas o notificaciones cuando se utiliza la IAGen, o mediante la creación de un sitio web de recursos públicos. Al crear un sitio web de este tipo, los gobiernos deben centrarse en publicar la mayor cantidad posible de información sobre los usos de la IAGen que afectan directamente a los ciudadanos; por ejemplo, si una agencia de investigación utiliza datos de los ciudadanos para perfeccionar un modelo de IAGen, si una organización de aplicación de la ley utiliza la IAGen en su labor policial o si un programa de red de seguridad social utiliza chatbots de IAGen para resumir las quejas de los ciudadanos.

Los sitios web de recursos públicos deben cubrir, entre otras áreas, dónde y cuándo se utiliza la IAGen, cualquier uso de datos personales en poder del gobierno o adquiridos por este, criterios de equidad y privacidad para adquirir productos de IAGen desarrollados comercialmente, controles y contrapesos para identificar y mitigar errores de IAGen y procesos internos como evaluaciones de impacto de la privacidad y mantenimiento de registros.

Si bien son sitios web en su etapa inicial y actualmente no incluyen muchos de estos componentes, la página web de la Oficina de Gestión de Personal de los Estados Unidos sobre la IAGen¹¹⁷ y el sitio web del estado de Nueva Jersey sobre el uso de la IAGen en el estado¹¹⁸ son dos enfoques para describir una colección de oportunidades, riesgos y usos de la IAGen de manera comprensible. Los gobiernos deberían publicitar sus sitios web educativos a los ciudadanos a través de campañas de comunicación pública, como en las plataformas de redes sociales. Tenga en cuenta que las OSC motivadas accederán principalmente a estos sitios web en lugar de los ciudadanos promedio, por lo que agregar etiquetas y otras notificaciones sobre el uso de la IAGen es primordial.

¹¹⁷ "Responsible Use of Generative Artificial Intelligence for the Federal Workforce."

¹¹⁸ "Using Generative AI in New Jersey State Government." State of New Jersey Office of Innovation, innovation.nj.gov/skills/ai. Accessed 26 June 2024.

Asegúrese de que los equipos de adquisiciones que adquieran o concedan licencias de soluciones de IAGen cuenten con un personal y un enfoque multidisciplinarios para garantizar el uso y la aplicación justos, basados en derechos y asuntos legales de la IAGen. Esto debería incluir la transparencia en los procesos y decisiones de adquisición. Por ejemplo, la ciudad de Seattle ha publicado una política¹¹⁹ que exige que todas las agencias gubernamentales pasen por el proceso de adquisición de tecnología de la información (TI) establecido en Seattle antes de usar cualquier software IAGen, software como servicio, servicio basado en la web, complemento de navegador o aplicación para teléfonos inteligentes y que sigan las pautas necesarias sobre privacidad, sesgo, daño, propiedad intelectual y rendición de cuentas: una política clara que es públicamente transparente.

Identificar oportunidades para integrar herramientas IAGen para aumentar la accesibilidad de los recursos gubernamentales y mejorar la prestación de servicios. Por ejemplo, considere dónde se puede integrar un chatbot impulsado por IAGen para responder preguntas de los electores,¹²⁰ mejorar la claridad del lenguaje¹²¹ en los sitios web y recursos gubernamentales, traducir documentos a más idiomas,¹²² proporcionar acceso a lectores de pantalla para el contenido, incluir más subtítulos en videos gubernamentales y anuncios de servicio público, y simplificar materiales escritos¹²³ para personas con bajo nivel de alfabetización. Las herramientas de IAGen también se pueden incorporar para facilitar el acceso a los servicios a través de mecanismos similares, como chatbots que pueden responder preguntas sobre los factores de calificación para recibir servicios y dirigir a los usuarios a páginas web relevantes y orientación sobre cómo solicitarlos. También puede volver a consultar los estudios de caso en las páginas 19 y 20 para ver cómo las herramientas de IAGen pueden generar ganancias de eficiencia. Y, para garantizar que las necesidades y los derechos de los ciudadanos sigan siendo centrales para cualquier integración de IAGen, asegúrese de que existan oportunidades de reparación y compensación para que los ciudadanos las utilicen cuando la integración de estas herramientas resulte en daños o consecuencias no deseadas.

Evaluar cómo la IAGen se adapta a las leyes y regulaciones ya vigentes. Esto incluye leyes de protección al consumidor o regulaciones de privacidad de datos; considere realizar investigaciones o publicar una guía regulatoria para productos y casos de uso específicos de la IAGen. Por ejemplo, consulte la investigación de la Superintendencia de Industria y Comercio de Colombia¹²⁴ sobre el cumplimiento de ChatGPT con la legislación de protección de datos del país y la de Estados Unidos,¹²⁵ para empresas sobre IA y Estados Unidos.

Si se identifican lagunas significativas en las leyes y regulaciones vigentes, se deben consultar las directrices existentes para iniciar un proceso legislativo o regulatorio que garantice que los sistemas de IAGen se utilicen de acuerdo con los principios democráticos y de derechos humanos. Entre las directrices útiles a las que se puede hacer referencia se encuentran los principios éticos de la IA de las Naciones Unidas antes mencionados, recursos como el Observatorio de Políticas de IA de la OCDE,¹²⁶ y regulaciones como la Ley de IA de la UE.¹²⁷

¹¹⁹ Balgetty, Ben. "City of Seattle Releases Generative Artificial Intelligence Policy Defining Responsible Use for City Employees." Office of Mayor Bruce Harrell, 3 Nov. 2023, harrell.seattle.gov/2023/11/03/city-of-seattle-releases-generative-artificial-intelligence-policy-defining-responsible-use-for-city-employees.

¹²⁰ Averyheart, Kathryn. "State and Local Agencies Can Improve the Citizen Experience With AI Customer Service." Technology Solutions That Drive Government, 15 Apr. 2024, statetechmagazine.com/article/2024/04/state-and-local-agencies-can-improve-citizen-experience-ai-customer-service.

¹²¹ Novek, Beth. "Generative AI and Policymaking for the New Frontier." GovTech, 1 Dec. 2023, www.govtech.com/opinion/generative-ai-and-policy-making-for-the-new-frontier.

¹²² Douglas, Theo. "State GenAI Opportunities Target Inspections, Language." Industry Insider, 1 Feb. 2024, insider.govtech.com/california/news/state-genai-opportunities-target-inspections-language.

¹²³ "Generative AI and Policymaking for the New Frontier."

¹²⁴ Gonzalez, Geussepppe, and Rodrigo Serrallonga Mejía. "Colombian Regulator Launches Inquiry Into ChatGPT Over Personal Data Concerns." Access Partnership, 16 May 2023, accesspartnership.com/access-alert-colombian-regulator-launches-inquiry-into-chatgpt-over-personal-data-concerns.

¹²⁵ Levine, Samuel. "Clear Eye on AI." FTC Business Blog, 10 Apr. 2024, www.ftc.gov/business-guidance/blog/2024/04/business-blogs-clear-eye-ai.

¹²⁶ "Policies, Data and Analysis for Trustworthy Artificial Intelligence." OECD Artificial Intelligence Policy Observatory, oecd.ai/en. Accessed 19 July 2024.

¹²⁷ "EU AI Act: First Regulation on Artificial Intelligence." European Parliament, 6 June 2023, www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

Los gobiernos podrían considerar posibles casos de uso, oportunidades y riesgos asociados con las campañas políticas que utilizan la IAGen para crear videos de campaña (por ejemplo, ¿está permitido y se requiere que las campañas etiqueten el contenido en consecuencia?¹²⁸), los ciudadanos que informan errores sobre ellos en los resultados de los chatbots (por ejemplo, ¿las empresas pueden y están obligadas a corregir las inexactitudes?¹²⁹), o las empresas que construyen modelos de IAGen con datos personales (por ejemplo, ¿qué significa el “consentimiento” del consumidor?¹³⁰), entre otros. Las leyes y regulaciones deben incluir principios de equidad, transparencia y rendición de cuentas al público.

Evaluar cómo el gobierno regula o audita las herramientas y acciones de la industria relacionadas con la IAGen.

Considere qué acciones puede tomar democráticamente el gobierno para proporcionar supervisión y proteger los derechos de los ciudadanos. Esto podría incluir agencias reguladoras que evalúen leyes y regulaciones ya vigentes para determinar si podrían aplicarse al desarrollo, implementación y uso de la IAGen y cómo, y emitir orientación general u opiniones consultivas formales para brindar claridad a las empresas y la sociedad civil sobre los casos de uso. Por ejemplo, la orientación pública de la Comisión Federal de Comercio de los Estados Unidos sobre las preocupaciones de competencia de la IAGen¹³¹ y cómo la ley de “engaño” de Estados Unidos se aplica a los términos de servicio de los modelos de IA¹³² son dos ejemplos de cómo evaluar las leyes y regulaciones existentes en relación con la IAGen y explicar las implicaciones claramente a las empresas y al público. India ofrece otro ejemplo, emitiendo un aviso¹³³ en diciembre de 2023 recordando a los intermediarios y plataformas las reglas de tecnología de la información existentes que impactan la IAGen y emitiendo otro aviso¹³⁴ en marzo de 2024 describiendo más detalladamente cómo esas mismas empresas deben usar la IAGen y modelos de lenguaje grande según la ley vigente. Otras opciones para los gobiernos incluyen establecer sistemas de auditoría de terceros para evaluar las herramientas de IAGen; realizar auditorías de primera parte de los mecanismos de gobernanza corporativa para los sistemas de IAGen, por ejemplo, aprovechando a las autoridades reguladoras existentes para que exijan el acceso a los controles de las empresas para evitar el uso indebido del modelo de IAGen; establecer multas escalonadas, similares al enfoque regulatorio basado en el riesgo de la Ley de Inteligencia Artificial de la UE,¹³⁵ para exigir a las empresas que rindan cuentas por comportamientos nocivos; y emitir solicitudes de información a las agencias gubernamentales que inviten a las empresas, la sociedad civil y los miembros del público a responder a preguntas relacionadas con la IAGen que luego podrían informar la futura elaboración de normas regulatorias de una agencia.

Involucrar a los ciudadanos para garantizar que sus comentarios se incluyan y se consideren en todas las etapas de la integración de la IAGen. Esto incluye crear oportunidades para que los ciudadanos opinen antes, durante y después de la integración de una herramienta IAGen. Esto se puede hacer a través de mecanismos como cabildos abiertos, formularios de comentarios virtuales, encuestas públicas e incluso pruebas de usuario una vez que un gobierno avanza con la integración de la herramienta. Invertir en aumentar la comprensión del gobierno sobre el sentimiento público en torno a la IAGen para predecir usos tanto aceptables como controvertidos de la IAGen. Organizaciones como Democracy X en Dinamarca han hecho un buen trabajo en el pasado para recopilar información de los ciudadanos¹³⁶ sobre cómo los tomadores de decisiones buscan aprovechar las soluciones de IA a través de sesiones de escucha y diálogos participativos. Ser intencional en la participación de los grupos más vulnerables o marginados cuyas voces a menudo son excluidas, como las mujeres y los grupos de derechos de las minorías.

¹²⁸ “Our Approach to Labeling AI-Generated Content and Manipulated Media.”

¹²⁹ “ChatGPT Provides False Information About People, and OpenAI Can’t Correct It.” Noyb, 29 Apr. 2024, noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it.

¹³⁰ Voica, Alexandru. “GenAI Apps Are Cloning Your Likeness Without Consent—and Might Make You Famous for All the Wrong Reasons.” Fortune, 11 Mar. 2024, fortune.com/2024/03/11/gen-ai-apps-cloning-likeness-consent-tech-artificial-intelligence-risk-law.

¹³¹ Staff in the Bureau of Competition & Office of Technology. “Generative AI Raises Competition Concerns.” FTC Office of Technology Blog, 29 June 2023, www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns.

¹³² By Staff in the Office of Technology and the Division of Privacy and Identity Protection. “AI (and Other) Companies: Quietly Changing Your Terms of Service Could Be Unfair or Deceptive.” FTC Office of Technology Blog, 13 Feb. 2024, www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/ai-other-companies-quietly-changing-your-terms-service-could-be-unfair-or-deceptive.

¹³³ Ministry of Electronics and IT. MeitY Issues Advisory to All Intermediaries to Comply With Existing IT Rules. 26 Dec. 2023, pib.gov.in/PressReleasePage.aspx?PRID=1990542.

¹³⁴ “MeitY Revises AI Advisory, Does Away with Government Permission Requirement – Update.” AZB & Partners, 21 Mar. 2024, <https://www.azbpartners.com/bank/meity-liberalizes-ai-advisory-dated-march-1-2024-following-industry-concerns-and-issues-revised-advisory-on-march-15-2024/>.

¹³⁵ “Article 99: Penalties.” EU Artificial Intelligence Act, artificialintelligenceact.eu/article/99. Accessed 26 June 2024.

¹³⁶ “Participation.” The Danish Board of Technology, tekno.dk/service/participation/?lang=en. Accessed 26 June 2024.

Contratar expertos externos para que ayuden con objetivos como la formación y la mejora de las competencias, sin depender de expertos externos y creando, con el tiempo, la capacidad, la experiencia y la dotación de personal internas necesarias. Los gobiernos deberían hacerlo reservando fondos para crear y publicar un aviso (o avisos) de empleo cuando sea posible para un experto (o expertos) en políticas de IA, o encargando a un empleado gubernamental actual, como un experto en políticas tecnológicas, la tarea de desarrollar un plan para aumentar la experiencia en IA en el gobierno.

Al mismo tiempo, los gobiernos deberían recurrir a las OSC a nivel local, regional y global que tienen experiencia en tecnología, medios, democracia e IA para explicar los temas de la IAGen a los funcionarios públicos y a los responsables de las políticas, y ayudar a desarrollar un plan para aumentar la experiencia gubernamental en IA durante los próximos años.

RECOMENDACIONES PARA LA INDUSTRIA

Las empresas del sector privado, incluidas las plataformas de redes sociales y los desarrolladores de IAGen, pueden adoptar las siguientes medidas para contribuir a garantizar que la IAGen se utilice de forma democrática dentro de sus propias organizaciones y en las organizaciones que compran o adquieren su tecnología, así como al servicio de los valores públicos y democráticos. Estas oportunidades de acción abarcan muchas fases y aspectos de la creación y el uso de IAGen, incluida la recopilación y el perfeccionamiento de los datos de entrenamiento, el diseño y el desarrollo de modelos, la implementación de IAGen, los controles contractuales y técnicos en torno al uso de la IAGen y las normas de las plataformas en línea sobre el manejo de contenido de IAGen.

Desarrollar la capacidad de los profesionales de la industria en relación con las implicancias éticas, legales, sociales y de derechos humanos de la IAGen. Considere la posibilidad de hacerlo mediante un mayor desarrollo de materiales educativos y una amplia difusión de los recursos existentes de la empresa, las prácticas de diseño de IAGen y las políticas de implementación de IAGen. Por ejemplo, los desarrolladores de IA como IBM¹³⁷ y Google,¹³⁸ las empresas consultoras como Deloitte¹³⁹ y PwC,¹⁴⁰ así como los investigadores académicos¹⁴¹ han publicado principios y directrices de diseño para la ética de la IA y la IAGen. Se puede crear un ciclo positivo cuando las empresas invierten intencionalmente en publicar información sobre sus propias prácticas, educan a los desarrolladores sobre las implicaciones de la IA y proporcionan recursos a las empresas y universidades que buscan capacitar a los desarrolladores en ética de la IA, creando en última instancia más incentivos para que otras empresas desarrollen y publiquen un conjunto de principios de diseño ético. Estos pasos también impulsan la rendición de cuentas, porque las OSC pueden utilizar la orientación de las empresas para identificar lagunas en las prácticas de la industria y exigir a empresas específicas que cumplan con los estándares. También es importante invertir en la participación de una diversidad de expertos que puedan ayudar a comprender los amplios impactos sociales, incluidos expertos en derechos digitales, académicos de los derechos de las mujeres y defensores de grupos marginados.

Proporcionar financiación sostenible y a largo plazo a la sociedad civil y a los socios académicos en el espacio más amplio de la IA. La industria tiene un papel que desempeñar para garantizar que todas las partes interesadas, especialmente las que están fuera del gobierno, puedan participar de manera significativa en las conversaciones en torno al uso democrático de la IAGen. Ofrecer recursos a grupos como la sociedad civil y el mundo académico, y especialmente a organizaciones de todo el Sur Global, puede contribuir en gran medida a garantizar un ecosistema de IAGen saludable. La financiación podría destinarse a apoyar actividades de investigación, promoción y capacitación que, en última instancia, probablemente ayuden a alentar a la industria a minimizar los daños que sus propias herramientas pueden plantear.

¹³⁷ "AI Design Ethics Overview." IBM, www.ibm.com/design/ai/ethics. Accessed 26 June 2024.

¹³⁸ "Our Principles." Google AI, ai.google/responsibility/principles. Accessed 26 June 2024.

¹³⁹ Guszczka, Josh, et al. "Human Values in the Loop." Deloitte Insights, 28 Jan. 2020, www2.deloitte.com/us/en/insights/focus/cognitive-technologies/design-principles-ethical-artificial-intelligence.html.

¹⁴⁰ Axente, Maria Luciana, and Ilana Golbin. "Ten Principles for Ethical AI." PwC Australia Digital Pulse, 12 Apr. 2022, www.pwc.com.au/digitalpulse/ten-principles-ethical-ai.html.

¹⁴¹ Brey, Philip, and Brandt Dainow. "Ethics by Design for Artificial Intelligence." AI And Ethics, Sept. 2023, <https://doi.org/10.1007/s43681-023-00330-4>.

A través de estas contribuciones, los líderes de la industria podrían tener un papel importante en la creación de un ecosistema de IA responsable y podrían ayudar a revertir los recientes recortes de financiación que han afectado al ecosistema de derechos digitales, fundamental para informar sobre el desarrollo responsable de IAGen.

Involucrar a los responsables de las políticas en sus procesos nacionales de elaboración de leyes sobre IA para garantizar que las opiniones de la industria se formen y luego se comuniquen al gobierno y a la sociedad civil. Estos esfuerzos pueden realizarse mediante la participación directa con los responsables de las políticas, así como mediante otros esfuerzos de múltiples partes interesadas. Por ejemplo, aunque no esté dirigido por la industria, el modelo sin fines de lucro Partnership on AI para este tipo de creación de consenso e intercambio de información reúne a organizaciones sin fines de lucro como la Asociación Estadounidense para el Avance de la Ciencia y Code for Africa, instituciones académicas como el Centro de IA en Tokio y la Universidad Nacional de Australia, y empresas como Google y Meta para discutir y promover las mejores prácticas en materia de equidad algorítmica, participación de múltiples partes interesadas, IA y elecciones, y más. Estos modelos sirven como un gran ejemplo para considerar emular en su propio contexto.

Publicar más información sobre el uso y las prácticas de desarrollo de IAGen, y compartir estándares éticos y de derechos humanos en sitios web y directamente con usuarios/clientes, periodistas y organizaciones de la sociedad civil. Las plataformas de redes sociales, los motores de búsqueda y otras empresas deberían tener una única página web o un conjunto de páginas web donde la IAGen y los usos de la IA se enumeren y expliquen de manera clara, sucinta y comprensible, junto con enlaces relevantes a formularios de exclusión o corrección de datos. Esto debería incluir una descripción de si la empresa recopila datos directamente de los usuarios/clientes para entrenar sistemas de IA y si la empresa compra o adquiere¹⁴² datos sobre usuarios/clientes de terceros (por ejemplo, comprando a corredores de datos, extrayendo datos de sitios web públicos) para entrenar sistemas de IA. Las empresas también deberían interactuar con las organizaciones de la sociedad civil sobre sus estándares éticos de IAGen y políticas de uso a través de presentaciones en conferencias como NeurIPS¹⁴⁴, una de las reuniones anuales de IA más grandes; debates en foros de múltiples partes interesadas como la Asociación sobre IA; y mediante la creación de redes de asesoramiento de la sociedad civil, similares al antiguo Consejo Asesor de Confianza y Seguridad de Twitter¹⁴³, donde puedan consultar y escuchar continuamente a la sociedad civil.

Desarrollar productos y servicios de acuerdo con los principios de derechos humanos, aprovechando para ello equipos multidisciplinarios. Esto incluye la creación de equipos de políticas de IA con abogados, ingenieros de productos y asesores externos expertos de la sociedad civil. Como punto de referencia, las Naciones Unidas establecen 10 principios para el uso ético de la IA:¹⁴⁴

- No causar daño;
- propósito definido, necesidad y proporcionalidad;
- seguridad y protección;
- equidad y no discriminación;
- sostenibilidad;
- derecho a la privacidad, protección de datos y gobernanza de datos;
- autonomía humana y supervisión;
- transparencia y explicabilidad;
- responsabilidad y rendición de cuentas; y
- inclusión y participación.

¹⁴² Burgess, Matt, and Reece Rogers. "How to Stop Your Data From Being Used to Train AI." WIRED, 10 Apr. 2024, www.wired.com/story/how-to-stop-your-data-from-being-used-to-train-ai.

¹⁴³ Zakrzewski, Cat, et al. "Twitter Dissolves Trust and Safety Council." Washington Post, 12 Dec. 2022, www.washingtonpost.com/technology/2022/12/12/musk-twitter-harass-joel-roth/.

¹⁴⁴ "Principles for the Ethical Use of Artificial Intelligence in the United Nations System." Accessed 26 June 2024.

Para implementar estos principios en el desarrollo y uso de IAGen, las empresas deberían comenzar por realizar una evaluación de impacto en los derechos humanos (el Instituto Danés de Derechos Humanos explica cómo empezar¹⁴⁵) y luego realizar evaluaciones actualizadas periódicamente para evaluar las deficiencias en las políticas y prácticas de ingeniería de la empresa. También deberían involucrar a grupos de la sociedad civil y expertos externos para realizar auditorías externas, crear y dotar de personal a juntas de revisión independientes o supervisar las salvaguardas no regulatorias para el diseño, la implementación y el uso éticos de IAGen, así como para garantizar que los productos y servicios se alineen con los principios de derechos humanos.

Crear intencionalmente espacios sostenidos y oportunidades recurrentes para compartir conocimientos. Es fundamental que los desarrolladores de herramientas mantengan oportunidades y esfuerzos sostenidos para conectarse con la sociedad civil a fin de garantizar que las perspectivas, inquietudes y aprendizajes de estos grupos se integren regularmente en el desarrollo de productos y los orienten. Estos espacios deben estar disponibles de manera continua más allá de la conclusión de una auditoría o después del lanzamiento de un producto; de esa manera, es más probable que las herramientas reflejen de manera integral las inquietudes y los valores de los grupos de la sociedad civil. Además, el intercambio de este conocimiento es una vía de doble sentido, por lo que es importante compartir el conocimiento técnico de los equipos técnicos para ayudar a los actores democráticos en el desarrollo de aplicaciones, capacitación y políticas en torno a la IAGen, como compartir la experiencia técnica con la sociedad civil sobre las mejores prácticas para el uso de IAGen en la traducción.

Por último, estos espacios de intercambio de conocimientos deberían incluir oportunidades para compartir conocimientos entre otras empresas que desarrollan o utilizan la IAGen, especialmente compartir desafíos y aprendizajes relacionados con casos de uso político y preocupaciones sobre la integridad de la información para informar un enfoque ecosistémico más holístico para estos desafíos.

Adaptar los esfuerzos de mitigación de daños a diversos entornos y necesidades de grupos de identidad—por ejemplo, brechas de accesibilidad y usabilidad de IAGen entre diferentes contextos y desafíos de países, o disparidades en cómo el discurso de odio y la desinformación afectan a las mujeres. Combatir la desinformación y el discurso de odio creados por IAGen en inglés en Europa occidental, por ejemplo, requerirá soluciones y enfoques diferentes a los que se utilizan para combatir la desinformación y el discurso de odio creados por IAGen en otro idioma (en el que muchos de los principales chatbots de IAGen no están entrenados.)¹⁴⁶

Destinar recursos a métodos para detectar el uso indebido de IAGen, como, por ejemplo, identificar la desinformación creada por IAGen¹⁴⁷ y otros contenidos¹⁴⁸ o identificar cuándo los individuos manipulan los chatbots de IAGen con mensajes¹⁴⁹ diseñados para devolver datos subyacentes de los usuarios. Colaborar con otras empresas y gobiernos en esos métodos de detección y asegurarse de que esas técnicas de detección se enumeren y expliquen a los usuarios/clientes en las páginas web mencionadas anteriormente para aumentar la transparencia.

¹⁴⁵ "Introduction to Human Rights Impact Assessment." The Danish Institute for Human Rights, www.humanrights.dk/tools/human-rights-impact-assessment-guidance-toolbox/introduction-human-rights-impact-assessment. Accessed 26 June 2024.

¹⁴⁶ North, Madeleine. "Generative AI Is Trained on Just a Few of the World's 7,000 Languages. Here's Why That's a Problem – and What's Being Done About It." WEF Foreign Agenda, 17 May 2024, www.weforum.org/agenda/2024/05/generative-ai-languages-llm.

¹⁴⁷ Loth, Alexander, et al. "Blessing or Curse? A Survey on the Impact of Generative AI on Fake News." arXiv, 3 Apr. 2024, doi.org/10.48550/arXiv.2404.03021.

¹⁴⁸ Perkins, Mike, et al. "GenAI Detection Tools, Adversarial Techniques and Implications for Inclusivity in Higher Education." arXiv, 28 Mar. 2024, doi.org/10.48550/arXiv.2403.19148.

¹⁴⁹ Iyer, Prithvi. "New Study Suggests ChatGPT Vulnerability With Potential Privacy Implications." Tech Policy Press, 29 Nov. 2023, www.techpolicy.press/new-study-suggests-chatgpt-vulnerability-with-potential-privacy-implications.



ADOPTAR LA COLABORACIÓN INTERSECTORIAL

Una forma de abordar la maximización de las oportunidades de la IAGen y la mitigación de sus riesgos es desglosar las recomendaciones por sociedad civil, gobierno e industria. Otra consideración importante, señalaron los miembros del grupo de trabajo, es el tipo de personas de una organización que implementan una recomendación, como abogados, ingenieros, formuladores de políticas o profesionales de la opinión pública. La transparencia es un gran ejemplo para ilustrar este punto; para los abogados, la transparencia podría significar comparar las prácticas de la empresa con una lista de verificación basada en regulaciones, con cada elemento de la lista abierto a cierta interpretación. Para un ingeniero, la transparencia podría significar la implementación de medidas técnicas específicas. Para un científico social, la transparencia podría significar la realización de un estudio cualitativo y cuantitativo de cómo interactúan los usuarios con los modelos y el contenido de la IAGen, y si consideran que los modelos o el contenido son confiables. Maximizar las oportunidades de la IAGen para la democracia y, al mismo tiempo, mitigar sus numerosos riesgos para la democracia requerirá este tipo de amplia participación de ingenieros de productos, científicos de datos, moderadores de contenido, formuladores de políticas, organismos de control de los medios y otros actores.

MIRANDO HACIA EL FUTURO: ¿QUÉ ANTICIPAR?

Dados los riesgos y las oportunidades que hemos planteado, ¿cómo podemos anticipar la evolución de los impactos de la IAGen? Si bien muchas de las recomendaciones del informe técnico tienen como objetivo equipar mejor a los actores democráticos para que comiencen a tomar medidas ahora para abordar los impactos de la IAGen, esta sección adopta una perspectiva prospectiva sobre cómo pueden evolucionar los impactos de la IAGen en los próximos años. Los miembros del grupo de trabajo destacaron una serie de posibles realidades inciertas que pueden concretarse dependiendo de cómo evolucione el espacio de la IAGen. La probabilidad de que se desarrollen los siguientes escenarios depende de factores que incluyen el alcance de la regulación de la IAGen; la eficacia de la regulación de la IAGen por parte de los gobiernos; la evolución de las herramientas de IAGen y las formas en que se vuelven cada vez más poderosas; y el grado en que la industria establece pautas de uso.

Si bien las realidades que se detallan a continuación aún están por verse, es importante que la sociedad civil, el gobierno, la industria y la academia consideren cómo se puede aplicar la IAGen en el futuro, y que la comunidad democrática continúe evaluando los impactos de IAGen.



RECOPIACIÓN DE DATOS

En ausencia de leyes y regulaciones sólidas sobre privacidad de datos, las empresas, los gobiernos y las instituciones de investigación podrían recopilar y utilizar datos personales (como medios protegidos por derechos de autor, imágenes faciales, grabaciones de voz y videos) en violación de la privacidad de los ciudadanos. Como resultado, los datos de los ciudadanos podrían usarse continuamente sin su consentimiento para desarrollar aplicaciones IAGen, lo que puede violar aún más su privacidad, por ejemplo, mediante la creación de imágenes deepfake no consentidas cada vez más precisas.



VIGILANCIA

En ausencia de una legislación protectora, los ciudadanos pueden estar sujetos a la vigilancia gubernamental para recopilar datos que permitan un mayor desarrollo de la IA, incluida, entre otras cosas, la recopilación de datos biométricos y la vigilancia del comportamiento en línea. Los gobiernos pueden aprovechar la IAGen para mejorar su capacidad de monitorear contenidos, censurar discursos y rastrear las actividades de disidentes y ciudadanos.



OPORTUNIDADES DE DIVERSIDAD LINGÜÍSTICA

Si los desarrolladores de IAGen invierten en contenido y datos de capacitación en idiomas distintos del inglés, podrían surgir oportunidades para que las organizaciones no gubernamentales (ONG) en defensa de la democracia, los organismos de control de la rendición de cuentas, las campañas políticas y las agencias gubernamentales de todo el mundo amplíen la accesibilidad de su contenido para incluir a hablantes de diferentes idiomas. Estas herramientas multilingües podrían mejorar los esfuerzos de integridad de la información en todas las plataformas en una diversidad más amplia de idiomas, incluidos los idiomas de las comunidades indígenas o étnicas que a menudo se han pasado por alto.



NORMAS INTERNACIONALES FORTALECIDAS O FRACTURADAS

La IAGen es sólo un componente de una conversación más amplia sobre los entornos de información, las democracias, las elecciones y las normas en relación con la tecnología. Si el rápido desarrollo y despliegue de los modelos de IAGen supera incluso las barreras de protección no reglamentarias de la IAGen más desarrolladas, el panorama de las normas internacionales puede fragmentarse aún más o alejarse aún más de los valores democráticos, incluidos, entre otros, la privacidad, la transparencia, la libertad de expresión, la rendición de cuentas y los derechos humanos.



LA FALTA DE EXPLICACIÓN DE LOS RESULTADOS DE AIGEN CONDUCE A UNA REDUCCIÓN DE LA RENDICIÓN DE CUENTAS

Debido a los persistentes y graves problemas de explicación en torno al funcionamiento de la IAGen, seguirá siendo difícil explicar cómo y por qué se crean determinados resultados. Si los gobiernos se apresuran a integrar la IAGen en las políticas públicas y los procesos de beneficios públicos, la falta de transparencia en cuanto a cómo y por qué los resultados de IAGen están informando las decisiones y los procesos gubernamentales reducirá la capacidad del público para exigir responsabilidades a esos procesos. Es posible que los gobiernos y los desarrolladores no puedan explicar por qué los sistemas IAGen llegaron a decisiones específicas. Los grupos de vigilancia de la sociedad civil, los medios de comunicación y los ciudadanos podrían pedir, presentar peticiones o demandas para tener acceso a más información sobre el uso y el desarrollo de la IAGen en esas situaciones, pero es posible que no tengan éxito.



DIVERSIFICACIÓN DEL ESPACIO DE DESARROLLO DE IA

Actualmente, el espacio de desarrollo de la IAGen está impulsado en gran medida por empresas con sede en Estados Unidos y China, pero es probable que más países implementen la IAGen, creando una base más diversificada para el desarrollo de IA. Este cambio puede dar como resultado herramientas más localizadas, lo que se traducirá en herramientas potencialmente más accesibles, pero que podrían fragmentar aún más el espacio de la IAGen. Sin embargo, la diversificación de las herramientas de IA dependerá de si los países adoptan políticas de datos abiertos y de la calidad y disponibilidad generales de los datos.



LA IMPORTANCIA CRECIENTE DE LAS HERRAMIENTAS SIN CÓDIGO

Las herramientas de IAGen sin código, como Hugging Face y muchas otras, aumentarán la capacidad de las personas con poca o ninguna experiencia en codificación para desarrollar herramientas de IAGen en plataformas de código abierto. Esto puede democratizar el acceso y el desarrollo, al mismo tiempo que satura el espacio de IAGen con herramientas de alta y baja calidad, lo que puede confundir a los usuarios y dar como resultado una plétora de servicios y productos no regulados con poca o ninguna supervisión.



CONFIANZA FRACTURADA EN EL ESPACIO DE LA INFORMACIÓN

A medida que la IAGen continúa avanzando en capacidades y el acceso a la IAGen se expande, la desinformación y las falsificaciones profundas desarrolladas por IAGen (y, como resultado, la conciencia y la sospecha sobre el contenido digital) seguirán creciendo, fracturando la confianza en la información en línea en su conjunto.

APÉNDICE: RECURSOS ADICIONALES

Hay muchos recursos disponibles para ayudar a comprender mejor la inteligencia artificial generativa, desde recursos educativos sobre terminología hasta artículos que hacen un seguimiento de su uso y sus impactos en todo el mundo. Consulte esta hoja de cálculo, que se actualizará continuamente, para obtener una lista cada vez mayor de recursos:

[Lista de recursos sobre inteligencia artificial generativa y democracia](#)



INTERNATIONAL REPUBLICAN INSTITUTE

IRI.org | @IRIglobal | P: (202) 408-9450 | E: info@iri.org