

Capítulo 14

El eje desinformación-inteligencia artificial en la estabilidad de la democracia*

DOI: <https://doi.org/10.25062/9786287818729.14>

María Eugenia Batista Cordeiro

Universidade Federal da Paraíba, Brasil

Vicente Torrijos

Escuela Superior de Guerra "General Rafael Reyes Prieto"

Resumen: Este capítulo examina el eje desinformación–inteligencia artificial como un factor crítico en la transformación de la estabilidad democrática en el siglo XXI. Se argumenta que la convergencia entre inteligencia artificial y ecosistemas digitales ha amplificado la capacidad de producir, personalizar y difundir desinformación a gran escala, alterando las condiciones del debate público y la formación de la opinión política. Desde un enfoque que integra la guerra cognitiva y la competencia informacional contemporánea, se analiza cómo la inteligencia artificial permite automatizar contenidos persuasivos, generar narrativas sintéticas (incluidos deepfakes) y microsegmentar audiencias, incrementando la eficacia de la manipulación informativa. Estas capacidades no solo expanden su alcance y velocidad, sino que dificultan su detección y mitigación por parte de Estados e instituciones democráticas. El capítulo sostiene que este fenómeno contribuye a la fragmentación de la esfera pública, la erosión de la confianza institucional y la polarización social, configurando un entorno donde la verdad factual pierde centralidad frente a narrativas emocionalmente resonantes. Asimismo, se examinan los riesgos de la automatización de la influencia política y la participación de actores estatales y no estatales en estas dinámicas. Finalmente, se plantea que la estabilidad democrática depende crecientemente del desarrollo de estrategias de defensa cognitiva que integren regulación tecnológica, alfabetización digital y capacidades avanzadas de detección y respuesta en entornos algorítmicos.

Palabras clave: democracia; desinformación; estabilidad; inteligencia artificial; regulación.

* Capítulo de libro resultado del proyecto de investigación "Desafíos y nuevos escenarios de la seguridad multi-dimensional a nivel nacional, regional y hemisférico en el decenio 2015-2025", del grupo de investigación Centro de Gravedad, de la ESDEG, categorizado A por MinCiencias y registrado con el código COL0104976. Los puntos de vista pertenecen a los autores y no necesariamente reflejan el pensamiento de las instituciones participantes.

María Eugenia Batista Cordeiro

Doctorada en Ciencias Políticas y Relaciones Internacionales en la Universidad Federal da Paraíba. Maestría en Derecho por la Universidad Federal de Rio Grande do Norte. Investigadora en el área de crimen organizado transnacional, flujos financieros ilícitos y otros. Miembro experto de la Global Initiative Against Organized Crime y egresado del George C. Marshall Center to European Studies y del W.J. Perry Center, National Defense University.

<https://orcid.org/0000-0002-1635-9013> - Contacto: maria.cordeiro@mprn.mp.br

Vicente Torrijos

Magíster en Estudios Políticos, Postgraduado en Altos Estudios Internacionales, estudios doctorales en Relaciones Internacionales y beca postdoctoral en Estudios Estratégicos, Seguridad y Defensa. Docente de la Escuela Superior de Guerra "General Rafael Reyes Prieto", y Profesor Adjunto del W.J. Perry Center, National Defense University.

<https://orcid.org/0000-0003-3837-6196> - Contacto: vicente.torrijos@esdeg.edu.co

Citación APA: Batista Cordeiro, M. E., & Torrijos, V. (2026). El eje desinformación-inteligencia artificial en la estabilidad de la democracia. En C. E. Álvarez Calderón (Ed.), *El campo de batalla digital: la defensa nacional en la era algorítmica, Volumen III. (Armas de la mente: guerra memética y manipulación informativa en la quinta generación de la guerra)* (pp. 1203-1222). Sello Editorial ESDEG.
<https://doi.org/10.25062/9786287818729.14>

EL CAMPO DE BATALLA DIGITAL: LA DEFENSA NACIONAL EN LA ERA ALGORÍTMICA

VOLUMEN III. ARMAS DE LA MENTE: GUERRA MEMÉTICA Y MANIPULACIÓN INFORMATIVA EN LA QUINTA GENERACIÓN DE LA GUERRA

ISBN IMPRESO (Obra completa): 978-628-7818-65-1

ISBN DIGITAL (Obra completa): 978-628-7818-69-9

ISBN IMPRESO Volumen I: 978-628-7818-68-2

ISBN DIGITAL Volumen I: 978-628-7818-72-9

DOI: <https://doi.org/10.25062/9786287818729>

Colección Seguridad y Defensa

Sello Editorial ESDEG

Escuela Superior de Guerra "General Rafael Reyes Prieto"

Bogotá D.C., Colombia

2026



Introducción

En estos tiempos de la revolución tecnológica de la información, las transformaciones en las formas de comunicación y relación han cambiado profundamente el comportamiento humano y las interacciones sociales. La inteligencia artificial (IA), producto emblemático de esta nueva era, ha surgido como una solución para una serie de problemas que afectan el entorno social, ejerciendo una influencia significativa en la ciencia, la educación, la comunicación y las relaciones sociales.

La IA se revela esencial tanto en la producción como en la diseminación de información. Sin embargo, a pesar de sus beneficios, también representa un desafío para las bases de los sistemas democráticos, pues tiene el potencial de propagar desinformación, erosionando la confianza de las poblaciones en sus instituciones democráticas y amenazando la estabilidad de estos sistemas.

Frente a este escenario, la inteligencia artificial se presenta como un instrumento de solución para las problemáticas que surgen, exigiendo un análisis de su capacidad para influenciar la dinámica democrática.

Este estudio busca comprender cómo la IA puede trascender esta visión de amenaza para convertirse en una herramienta en el combate contra la desinformación, reforzando así la estabilidad de las democracias modernas.

Esta investigación plantea la pregunta de si la inteligencia artificial puede ser efectivamente utilizada para disminuir la desinformación y, consecuentemente, fortalecer las democracias contemporáneas. La hipótesis orientadora sugiere que, mediante la aplicación de mecanismos éticos y regulatorios adecuados, especialmente de transparencia y democracia algorítmica, la inteligencia artificial puede disminuir la propagación de desinformación, así como mejorar la calidad y veracidad de las informaciones disponibles al público. Para probar esta hipótesis, se adoptará el método hipotético-deductivo, implementado por medio de una

revisión bibliográfica sistemática, utilizando bases de datos renombradas, como *Web of Science*.

Mediante un levantamiento bibliográfico en el indexador *Web of Science - Colección Principal (Clarivate Analytics)*, enfocándose en las intersecciones de las palabras clave *democracy*, *artificial intelligence*, *fake news* y *disinformation*, se localizaron 963 artículos, todos producidos durante 2023. De los artículos localizados, fueron seleccionados los más referenciados, siendo utilizados aquellos que más se adecuaron a la temática. Adicionalmente, se incorporaron regulaciones recientes de la Unión Europea relacionadas con la desinformación y la inteligencia artificial, destacando el *Código de buenas prácticas contra la desinformación* y la nueva *Ley de Inteligencia Artificial*. La reciente resolución de la Asamblea General de la ONU también fue considerada para contextualizar los desafíos éticos y regulatorios que se enfrentan actualmente.

Además, los pensamientos de Jürgen Habermas, John Rawls y Yochai Benkler también fueron utilizados como fundamento teórico destinado a la comprensión de las dinámicas sociales que involucran conceptos de esfera pública, participación democrática, opinión pública y consenso, cuyo enfoque determinado fue la influencia de la IA en combate y propagación de la desinformación.

El presente artículo está estructurado en tres tópicos principales, además de los elementos introductorios y conclusivos. El primero de ellos, titulado "Democracia en la era digital", investiga cómo la revolución digital ha transformado el ejercicio democrático mediante la remodelación de la gobernanza y el compromiso cívico. Esta parte del artículo discute la capacidad de la tecnología para alterar radicalmente las estructuras sociales existentes, destacando cómo estos cambios afectan la interacción entre ciudadanos y gobiernos. Se examinará cómo las herramientas digitales han permitido un diálogo más amplio e inclusivo entre los gobernantes y la población, potencializando la participación cívica en procesos políticos anteriormente limitados por barreras geográficas o burocráticas.

Esta sección también aborda los desafíos inherentes a la era de la información, como la dificultad de verificar la autenticidad de la información y gestionar el exceso de contenido disponible. El papel de las redes sociales se discute críticamente, considerando su impacto significativo en la modulación de la opinión pública y en la manipulación política. Se analizará cómo las plataformas de medios sociales, al democratizar el acceso a la publicación de contenido, también han creado nuevas oportunidades para la propagación de informaciones falsas o engañosas que pueden desestabilizar los procesos democráticos.

En la segunda sección, “Democracia, inteligencia artificial y desinformación”, explora la dualidad de la inteligencia artificial en el contexto político. Por un lado, la IA ofrece herramientas poderosas para analizar grandes volúmenes de datos, posibilitando nuevas formas de compromiso cívico y gobernanza transparente. Por otro lado, la misma tecnología puede ser utilizada para crear y diseminar desinformación, representando un riesgo significativo para la estabilidad democrática. Esta parte del artículo discutirá la definición y los impactos de la desinformación en los regímenes democráticos, enfatizando cómo la desinformación puede minar la confianza pública en las instituciones e influir adversamente en los resultados electorales y las políticas públicas. Se examinarán las estrategias mediante las cuales la IA es empleada para identificar y filtrar contenido falso, así como las limitaciones y desafíos éticos asociados a estas tecnologías. Los análisis incluirán estudios de caso relevantes que ilustran tanto los éxitos como los fracasos en el uso de la IA para combatir la desinformación.

Por último, “Desafíos éticos y legales de la IA en el combate a la desinformación” es la temática de la última sección y tratará del uso de IA para detectar desinformación, abordando las complejidades involucradas en el empleo de la IA dentro del contexto democrático. La discusión se centra en los dilemas éticos y legales que surgen al utilizar algoritmos avanzados para detectar y combatir la desinformación. Se explorará la necesidad de garantizar que los derechos individuales no sean comprometidos por la vigilancia o censura automatizada. También se discutirá cómo los principios de transparencia, responsabilidad y derechos humanos pueden ser integrados en la concepción e implementación de tecnologías de IA. La aproximación regulatoria y la legislación existente serán analizadas para entender cómo diferentes jurisdicciones están respondiendo a los desafíos presentados por la IA.

En la última parte, se pretende entender los desafíos actuales enfrentados por las democracias en la era digital y, también, explorar caminos tecnológicos como soluciones posibles para el fortalecimiento de las bases democráticas.

Democracia en la era digital

Esta sección explora las transformaciones en el ejercicio democrático impulsadas por la revolución digital, evaluando cómo esta ha remodelado la gobernanza y el compromiso cívico.

La tecnología ha demostrado una capacidad sin precedentes para alterar las estructuras sociales existentes, modificando la interacción entre los ciudadanos y los gobiernos establecidos. Un análisis detallado revela no solo los avances, sino también los desafíos presentados por la era de la información, incluida la dificultad de gestionar la veracidad y el exceso de información, así como el impacto de estos factores en el debate público. Además, el papel de las redes sociales, que se han convertido en herramientas poderosas de influencia y manipulación política, es un punto crucial de discusión, especialmente considerando cómo estos medios afectan la modulación de la opinión pública.

En el contexto más amplio de las prácticas comunicacionales democráticas, Benkler (2006) destaca características fundamentales para un sistema de comunicación efectivo en una democracia, que incluyen: 1) captación universal, una esfera pública con la habilidad de capturar las observaciones y opiniones de todos los constituyentes, asegurando que cualquier persona crea que sus consideraciones serán escuchadas y tenidas en cuenta; 2) filtración para relevancia política potencial, separación de asuntos dentro del dominio de la acción política organizada, con enfoque en temas relevantes para la acción colectiva; 3) filtración para acreditación, que se relaciona con la credibilidad de las declaraciones en la esfera pública, de modo que sean verosímiles y relevantes para el debate político; 4) síntesis de la "opinión pública", el sistema de comunicación permite la síntesis de opiniones individuales en posiciones articuladas, siendo estas consideradas en la esfera política formal; y, además, 5) independencia del control gubernamental, la esfera pública política debe permitir que las observaciones de individuos y la opinión pública estén libres del control del gobierno.

Desde una concepción formulada como pluralismo razonable, basado en la admisión de doctrinas divergentes, resultantes del racionalismo y la libertad humana en una sociedad democrática, John Rawls (1993) presenta la idea de consenso superpuesto, que tiene como horizonte la concepción política de la justicia como equidad.

Así, para Rawls (1993), la idea de la justicia como equidad se basa en la capacidad de una sociedad para lograr un consenso sobre cuestiones fundamentales para la vida social en su entorno democrático.

Para el desarrollo de esta idea de justicia como equidad, es fundamental que la sociedad sea vista como un sistema equitativo de cooperación, que tiene como presupuestos: 1) la cooperación se basa en reglas aceptadas voluntariamente, en lugar de actividades coordinadas por órdenes de una autoridad; 2) los términos

de la cooperación deben ser justos y basados en la reciprocidad, asegurando que todos los que contribuyen reciban beneficios proporcionales; 3) la cooperación busca alcanzar el bien o la ventaja racional de todos los involucrados, según lo definido por ellos mismos (Rawls, 1993).

El consenso sobre estas cuestiones fundamentales para la vida en sociedad en un sistema democrático pasa por la construcción de lo que Jürgen Habermas (1991) definió como esfera pública, que se traduce en un espacio en el cual la población desarrolla la comunicación libre e inclusiva, con autonomía para expresar sus opiniones, comprometidas en un diálogo para alcanzar un consenso, mediante el debate público.

En una sociedad cada vez más integrada, los nuevos medios de comunicación crean oportunidades para el desarrollo de estas esferas públicas, con mayor potencial de inclusión y participación popular, resultando en estructuras más justas e igualitarias.

Después de todo, internet y la economía de la información basada en redes ofrecen una plataforma revolucionaria para la esfera pública, que se mantiene autónoma tanto del control gubernamental como de las presiones del mercado, formando un escenario que favorece la colaboración y el compromiso de los ciudadanos en procesos de producción y discusión colectiva (Benkler, 2006). En este contexto, la producción y circulación de información en grandes volúmenes de datos, principalmente a través de las redes sociales, influye directamente en la esfera pública y la opinión colectiva, transformando la democracia en un escenario más dinámico e informado.

Como bien ha observado Benkler (2006), internet genera diferentes impactos en la esfera pública en relación con los regímenes políticos, aunque también afecta significativamente la práctica cultural y la comunicación en todos los contextos. También señala el autor que, en regímenes autoritarios, internet crea un desafío mayor a los gobiernos en gestionar la esfera pública mediante el control del flujo de información en la sociedad. Por otro lado, en regímenes democráticos, internet tiene el don de influir en aspectos económicos y organizacionales de la sociedad, transformando la esfera pública en un ambiente de diálogo continuo, con mayor participación social (Benkler, 2006).

Como producto de esta nueva era de la información, las redes sociales emergen en el debate democrático influenciando y moldeando profundamente la opinión de los receptores e influyendo en la formulación del consenso en diversas esferas del conocimiento, especialmente en aquellos dominios que benefician a

actores con alto poder financiero, los cuales disponen de mayor capacidad para modular la opinión pública en favor de sus intereses privados.

Aunque la tecnología ha revolucionado los medios de comunicación, con potencial para el desarrollo de una esfera pública más inclusiva, se ha observado un movimiento contrario a estas nuevas plataformas de interacción social, que afectan la legitimidad de esos debates.

En un estudio sobre comportamientos en el ciberespacio, publicado por el sociólogo Sergio Amadeu da Silveira, se explora la relación entre sistemas digitales algorítmicos y el debate democrático, destacando cómo los algoritmos actúan en la modulación de la opinión pública. Los algoritmos se describen como secuencias lógicas de rutinas que procesan información. Además, la combinación de bases de datos, modelos matemáticos y software es esencial en la formación de la opinión pública y en la influencia sobre preferencias políticas de los votantes (Silveira, 2019).

En este contexto, es importante mencionar el papel de los bots en las plataformas de redes sociales, especialmente ante la capacidad que estos tienen de crear cuentas y transformarlas en perfiles influyentes. Estos robots, tras reunir un número considerable de seguidores, están programados para desinformar y lanzar ataques contra instituciones democráticas y adversarios políticos, alcanzando a menudo resonancia en los temas de tendencia (Castro & Cordeiro, 2022). Para ello, abusan de las estrictas reglas de privacidad inherentes a estas plataformas.

Esta dinámica deslegitima la idea de un pluralismo razonable, necesario para la construcción de un consenso superpuesto, defendido por Rawls (1993) y amenaza los cimientos sobre los que se fundan los sistemas democráticos. Como bien ha explicado Benkler (2018):

Estaba ocurriendo algo importante que amenazaba la democracia y nuestra mirada colectiva se dirigía hacia lo nuevo y rápidamente cambiante: la tecnología. Procesos tecnológicos fuera del control de cualquier persona o país —la convergencia de redes sociales, la curaduría algorítmica de noticias, bots, inteligencia artificial y análisis de grandes datos— estaban creando cámaras de eco que reforzaban nuestros sesgos, eliminaban los indicios de confiabilidad y, en general, nublaban nuestra capacidad de dar sentido al mundo y, con ello, nuestra capacidad de gobernarnos como democracias razonables. (p. 4)

Estas realidades tecnológicas, creadas mediante inteligencia artificial, se apoyan en sesgos cognitivos y representan un peligro real para los sistemas

democráticos. Las amenazas por el uso malintencionado de la IA imponen desafíos que incluyen la detección de contenido digital realista, la atribución de fuentes, la obtención de explicabilidad y la incorporación de contexto y técnicas de fusión en análisis forenses, lo que demanda la colaboración entre los gobiernos, la sociedad y la academia (Cardenuto et al., 2023).

En el escenario que se perfila, la revolución digital ha demostrado tanto su potencial emancipador como sus peligros, reformulando las bases del ejercicio democrático y de la participación cívica. La capacidad de las nuevas tecnologías para remodelar la esfera pública e influir directamente en las dinámicas de poder exige la necesidad de enfoques regulatorios y éticos más robustos.

A medida que se transita hacia el próximo tema, "Democracia, inteligencia artificial y desinformación", es crucial examinar cómo la IA, que ha sido una herramienta de gran influencia en las transformaciones discutidas, también desempeña un papel ambivalente al servir tanto como vehículo para la desinformación como medio potencial para combatirla. Este dualismo despierta cuestionamientos no solo a las capacidades técnicas de la IA, sino también a las implicaciones éticas y sociales de su uso en las democracias liberales.

Democracia, inteligencia artificial y desinformación

Habiendo establecido cómo la era digital ha transformado la gobernanza democrática, el siguiente paso es entender el doble papel de la inteligencia artificial en la difusión y combate de la desinformación. Este tema profundizará en el análisis de cómo la IA puede simultáneamente desafiar y proteger los fundamentos de las democracias modernas.

El cambio de la arquitectura de medios de masa centralizada a una arquitectura en red, ha destruido las barreras de costo para la participación en la esfera pública. La transformación tecnológica que actualmente presenciamos está fundamentada en la formación de conexiones multilaterales, en tiempo real y a costo reducido, transformando a los individuos de meros receptores a sujetos activos del discurso público (Benkler, 2006).

Manuel Castells argumenta que, tradicionalmente, los medios como sistema de comunicación operan solo en una vía, que es la del transmisor al receptor. Sin embargo, una verdadera comunicación exige interacción e interpretación de mensajes entre estos dos actores (Castells, 2019), algo facilitado por las nuevas tecnologías. Las redes sociales, en particular, han revolucionado la forma de acceder

y de compartir información, abarcando desde noticias periodísticas hasta expresiones de opiniones y de grupos específicos.

Además, si por un lado estos nuevos canales de comunicación han facilitado la difusión de marketing alineado con estrategias comerciales, por otro lado, también se han convertido en un terreno fértil para la propagación de desinformaciones. Este tipo de contenido —que es intencionadamente falso o inexacto, creado para engañar e inducir al error— puede tener objetivos lícitos o ilícitos en favor de poseedores de poder económico, entre los cuales se incluye la influencia de elecciones o campañas de *lobby*.

Para los fines de esta investigación, se emplea el término *desinformación* como una categoría abarcadora destinada a incluir todas las formas de información falsa o inexacta, las cuales son deliberadamente creadas para inducir al público objetivo a adoptar una percepción o comprensión distorsionadas sobre temas específicos (Adams et al., 2023).

La desinformación en las redes sociales alcanza una amplia audiencia, poseyendo un potencial de diseminación superior al de las noticias verídicas y captando particularmente la atención de individuos involucrados en debates políticos. A pesar de la existencia de servicios de verificación de hechos y metodologías y manuales de identificación para contrarrestar las noticias falsas, estas aproximaciones enfrentan obstáculos significativos, debido a la diversidad de contenidos y estándares encontrados en las noticias, lo que complica la gestión del gran volumen de información falsa que circula en línea (Kondamudi et al., 2023).

Las consecuencias de este fenómeno alcanzan las esferas social e individual, Adams et al. (2023) categorizan los efectos de la desinformación en estos aspectos. En el escenario social, la desinformación tiene su impacto: 1) en los medios, al deteriorar la confianza, cuestionando la verosimilitud de los medios tradicionales de información, así: 2) en la política, debilitando la democracia e influyendo en las creencias públicas y en el apoyo a las mismas; 3) en la ciencia, lo que afecta el conocimiento público, particularmente en la comunicación sobre salud y cambios climáticos, llevando a equívocos que impiden una acción política; y 4) en la economía, la desinformación es fuente de perturbación de mercados, influenciando el comportamiento del consumidor y la reputación de marcas. A nivel individual, la desinformación tiene efectos psicológicos, cognitivos y comportamentales (Adams et al., 2023).

Recientemente, el Foro Económico Mundial (WEF) publicó el *Informe de Riesgos Globales 2024* que señaló a la desinformación como el riesgo global más

grave previsto para los próximos dos años, quedando detrás de categorías como eventos climáticos extremos, inseguridad cibernética, conflictos armados, migración involuntaria, recesión económica, entre otros. Además, la polarización social figura como uno de los tres principales riesgos en el mismo horizonte temporal (WEF, 2024).

En ambientes de elecciones sistemáticas, la desinformación ha sido usada como un triunfo por los oponentes, ya que la diseminación de contenido falso pone en jaque tanto el proceso electoral como la legitimidad de los gobiernos elegidos, causando inestabilidad social. Como también señala el *Informe de Riesgos Globales* (WEF, 2014), este proceso de erosión de la verdad tiene un impacto directo en la polarización y en el discurso público, afectando el consenso en áreas relevantes como salud pública y justicia social, creando un ambiente propicio para la propagación de regímenes autoritarios que utilizan estos medios para llegar al poder, mantenerse en él y controlar su población:

En el contexto de la detección de desinformación, técnicas avanzadas como las redes neuronales convolucionales (CNNs) y redes neuronales recurrentes (RNNs), se utilizan para enfrentar los desafíos impuestos por las noticias falsas en plataformas de medios sociales. La aplicación de métodos de ensamble, que combinan diversas técnicas para mejorar la precisión y la robustez en la identificación de información falsa, ejemplifica cómo los principios de robustez y responsabilidad pueden ser aplicados en la práctica. (Tufchi et al., 2023, p. 10)

Muchas son las posibilidades traídas por la IA que pueden ser usadas para esparcir o identificar productos de desinformación, considerando que estas aplicaciones están basadas en *Training Data*, es decir, el desarrollo de estas se construye mediante el entrenamiento de algoritmos, cuya interacción puede ser supervisada o no, como se explica a continuación:

En la IA “supervisada”, los puntos de datos individuales se categorizan y etiquetan para que los algoritmos los puedan leer; luego, los algoritmos clasifican los datos e intentan clasificarlos correctamente según las etiquetas pre-determinadas. Los desarrolladores pueden dar retroalimentación directa a los modelos evaluando la calidad de sus clasificaciones. Repitiendo este proceso una y otra vez, los modelos de IA pueden “aprender” a identificar correctamente puntos de datos no etiquetados fuera del conjunto de datos original. (Neaher et al., 2024, p. 8)

En la IA “no supervisada”, los algoritmos clasifican los datos por su cuenta, utilizando análisis estadístico avanzado para encontrar similitudes o conexiones entre puntos de datos. Este es el método utilizado por modelos de IA generativos como el ChatGPT, que crea respuestas a consultas buscando patrones entre vastas cantidades de datos de entrenamiento. Los desarrolladores dan retroalimentación sobre los resultados producidos por las IA, aunque tienen poco conocimiento sobre cómo la IA obtiene esos resultados (Neaher et al., 2024).

Aunque los avances proporcionados por la tecnología deben ser celebrados, es necesario reconocer los desafíos que se presentan, especialmente en relación con la desinformación. La transición hacia un escenario cada vez más interactivo y multilateral, donde las fronteras entre transmisores y receptores se vuelven cada vez más inmediatas, impone la urgencia de enfoques éticos y legales robustos en el uso de la IA.

Por todo lo anterior, es necesario garantizar que las tecnologías de detección de noticias falsas no solo identifiquen la desinformación con precisión, sino que también operen bajo reglamentaciones orientadas por los derechos humanos, imponiendo a los actores involucrados responsabilidades, la sujeción de sus decisiones tecnológicas a la transparencia y al escrutinio público. La convergencia de principios favorece la construcción de un debate público más informado y equitativo, destacando los esfuerzos para crear un ambiente donde la tecnología amplíe las capacidades humanas y promueva una inclusión más amplia.

Desafíos éticos y legales de la IA en el combate a la desinformación

Comprender el impacto de la IA en la propagación y detección de desinformación nos lleva naturalmente a los complejos desafíos éticos y legales involucrados. Esta sección se centra en discutir cómo las prácticas regulatorias y éticas pueden moldear el uso futuro de la IA, con el objetivo de fortalecer las democracias frente a las amenazas de la desinformación.

La implementación de los principios de transparencia, responsabilidad y derechos humanos en el uso de la IA para combatir la desinformación puede constituirse en un instrumento eficaz para el fortalecimiento de los sistemas democráticos. Al emplear la tecnología alineada con los valores éticos y democráticos, se contribuye a la promoción de sociedades más justas y equitativas. Así, la

tecnología puede establecerse como un vector de refuerzo para los pilares democráticos esenciales.

Como se afirmó en páginas anteriores, Benkler (2006) enumera los principios fundamentales que caracterizan a la comunicación efectiva en una sociedad, capaz de escuchar las opiniones de los individuos que la componen, de filtrar contenidos creíbles y de relevancia política, de analizar y articular la opinión pública con independencia del control gubernamental. Estos elementos de una comunicación efectiva, que favorecen la construcción de un espacio público inclusivo, son fundamentales en un sistema democrático, ya que promueven la acción colectiva y el debate político.

En un espacio público inclusivo, la transparencia es esencial para la credibilidad de las declaraciones en ese espacio, ya que la transparencia debe ser un principio central para el desarrollo de una IA confiable, especialmente ante el poder que tienen los sistemas basados en esta tecnología para moldear el debate público e interferir en los procesos políticos.

Los desafíos impuestos por la IA atraviesan la incorporación de los principios de transparencia, explicabilidad y responsabilidad, que encuentran barreras en la confidencialidad de las empresas sobre sus algoritmos y métodos. Además, los riesgos de la ciberseguridad, el sesgo y discriminación incorporados en los datos y algoritmos, violaciones de propiedad intelectual, preocupaciones con la privacidad, impactos sociales adversos como el desempleo y desafíos de sostenibilidad ambiental debido al alto consumo de energía, son todos riesgos significativos asociados a la IA. También hay debates sobre los impactos nocivos no intencionales a largo plazo de la IA, subrayando la necesidad de prácticas responsables de IA para mitigar posibles consecuencias negativas (Neaher et al., 2024).

El presente estudio parte del supuesto de que el cumplimiento de principios éticos en el uso de la IA es crucial y, por ello, están incorporados en varias iniciativas globales que pretenden lanzar el debate sobre el uso benéfico de esta tecnología.

Una de estas iniciativas partió de la Organización para la Cooperación y el Desarrollo Económico (OCDE), que desde 2019 desarrolla estándares y principios que deben ser adoptados por sus países miembros y socios estratégicos, enfatizando la necesidad del desarrollo de una IA que sea confiable y esté alineada con los derechos humanos y los valores democráticos (OCDE, 2021). Para ello, desarrolló un concepto de IA confiable, definido así:

IA confiable se refiere a sistemas de IA que incorporan los Principios de IA de la OCDE, es decir, sistemas de IA que respetan los derechos humanos y la

privacidad; son justos, transparentes, explicables, robustos, seguros; y los actores involucrados en su desarrollo y uso deben ser responsables [...] La OCDE (2021) establece cinco principios basados en valores para la administración responsable de la IA confiable: Crecimiento inclusivo, desarrollo sostenible y bienestar [...] Valores centrados en el ser humano y justicia [...] Transparencia y explicabilidad [...] Robustez y seguridad; Responsabilidad [...] Los Principios de la IA de la OCDE también contienen cinco recomendaciones para políticas nacionales y cooperación internacional. Las recomendaciones incluyen: 1) invertir en investigación y desarrollo de IA; 2) fomentar un ecosistema digital para IA; 3) moldear un entorno político propicio para IA; 4) construir capacidad humana y preparar para la transformación del mercado de trabajo; y 5) cooperación internacional para la IA confiable.

Estos principios no limitan la innovación de esta tecnología. Al contrario, proponen una innovación que promueva la expansión de las capacidades humanas y la inclusión de poblaciones marginadas, reconociendo el potencial que tiene la IA en reducir las desigualdades económicas, sociales, de género y ambientales, fortaleciendo los pilares de la democracia (OCDE, s.f.).

En cuanto a los desafíos impuestos por la desinformación, los principios de la OCDE alientan la adopción de responsabilidad y transparencia en el uso de IA, asegurando que los actores responsables por el desarrollo de estas herramientas se comprometan con el suministro de información precisa y comprensible sobre las fuentes de datos y los procesos que conducen a las decisiones, predicciones y resultados. La transparencia permite que los individuos impactados por esta tecnología puedan entender y cuestionar su producto. Además, la transparencia refuerza la confianza pública y asegura que el uso de la IA en contextos democráticos sea justo y accesible a un minucioso escrutinio (OCDE, s.f.).

En cuanto a seguridad y robustez, los principios de la OCDE (s.f.) refuerzan la necesidad de que los sistemas de IA operen de forma segura y confiable durante todo su ciclo de vida, incluyendo mecanismos para evitar comportamientos indeseables o potencialmente dañinos de estos sistemas. Además, las herramientas de IA deben emplearse con un compromiso a la integridad de la información, con el fin de evitar la influencia negativa en los procesos democráticos (OCDE, s.f.).

Finalmente, como aspecto central de estos principios, se exige que los actores involucrados en el desarrollo de sistemas de IA garanticen tanto la funcionalidad adecuada de los sistemas, como el cumplimiento de principios éticos y regulatorios a lo largo de todo el ciclo de vida de la IA. Para ello, se establece una gestión de riesgos que permita rastrear todas las decisiones y procesos (OCDE, s.f.).

Las iniciativas regulatorias alrededor del mundo pretenden asegurar que el desarrollo de aplicaciones de inteligencia respeten los principios éticos y los derechos humanos fundamentales. Puesto que la IA supervisada tiene la capacidad de ser programada tanto para honrar estos valores como para violarlos, el marco regulatorio permite la orientación y supervisión sobre la implementación de los principios establecidos.

En el ámbito de la Unión Europea (2023) se han creado varios acuerdos y normas legales para regular la temática. Algunos de ellos son: 1) Código de buenas prácticas en materia de desinformación, que establece un compromiso voluntario de plataformas en línea, empresas de medios y publicidad en la lucha contra la desinformación; 2) Ley de datos, que tiene la misión de proteger los datos personales, creando reglas para el uso de la información; 3) Ley de servicios digitales, que regula los mercados y servicios digitales, con el fin de proteger los derechos individuales; 4) Ley de gobernanza de datos, que regula el proceso de compartir datos, implementando un marco de gobernanza que promueva la confianza en el uso de los mismos.

Además, existe la *Ley de Inteligencia Artificial* de la Unión Europea, un reglamento que establece normas obligatorias para los desarrolladores de estas aplicaciones, garantizan la seguridad y la ética de estos sistemas. Por medio de esa ley, se crea una categorización de riesgos y se establecen los requisitos que los dueños de estos sistemas deben cumplir rigurosamente. Además, el documento establece el respeto a los valores fundamentales de la Unión Europea, como la protección de los derechos humanos, la democracia y el Estado de derecho (Parlamento Europeo, 2024).

En marzo de 2024, la Asamblea General de las Naciones Unidas presentó un proyecto de resolución, destacando el papel de la IA para alcanzar los objetivos de la Agenda 2030, relacionados con el desarrollo sostenible. El documento enfatiza la necesidad de que estos sistemas tengan como centro a la persona humana y a principios éticos, preservando los derechos humanos y la privacidad a lo largo de todo el ciclo de vida de la IA. Igualmente, resalta los riesgos de un diseño inadecuado de IA, como el aumento de las desigualdades digitales y la violación de los derechos humanos. Para evitar estos riesgos, la resolución propone la formación de un consenso global y la cooperación internacional para desarrollar normas seguras, motivando a los países en desarrollo a superar las divisiones digitales y promover una participación efectiva en la gobernanza de la IA (United Nations General Assembly, 2024).

El llamado a los países en desarrollo para colaborar en la formación de un consenso global sobre la temática no es casual. La distribución de la participación en el desarrollo de principios, estándares y regulaciones de IA segura y responsable es bastante desigual, ya que el 63 % de los principios de IA provienen de Europa, Estados Unidos y China; mientras que solo el 2 % se origina en países africanos; 5 % en América Latina y el Caribe y 19 % en Asia. Esta disparidad representa un riesgo significativo, pues las normas y reglas creadas en el Norte Global no siempre reflejan adecuadamente las diferencias políticas, sociales, económicas y culturales del Sur Global (Instituto Igarapé, 2024).

En última instancia, la resolución propuesta por la Asamblea General de la ONU en 2024 subraya la importancia de alinear la inteligencia artificial (IA) con los objetivos de desarrollo sostenible. Se resalta la necesidad de integrar principios de transparencia, responsabilidad y respeto a los derechos humanos en la implementación de la IA, enfatizando el papel vital de una participación equitativa en la formulación de normas globales. Especialmente crucial para los países en desarrollo, este enfoque asegura que las regulaciones de IA promuevan una gobernanza que respete la diversidad global y fortalezca las estructuras democráticas y sociales esenciales.

Conclusiones

Ante las complejas interacciones entre la inteligencia artificial (IA), la democracia y la desinformación, se comprende que, aunque la IA representa una revolución tecnológica con potencial para promover avances significativos en diversas esferas sociales, también presenta desafíos significativos. Estos desafíos incluyen la propagación de desinformación, que puede corroer la confianza pública en las instituciones democráticas e influir negativamente en las percepciones y comportamientos de los ciudadanos, amenazando la estabilidad democrática.

La implementación de principios éticos y regulatorios rigurosos es fundamental para que la inteligencia artificial pueda efectivamente combatir la desinformación y reforzar las democracias contemporáneas. La gobernanza de la IA que promueva transparencia, responsabilidad y protección de datos es esencial para mitigar los riesgos asociados a su uso indebido. Las regulaciones deben asegurar que la tecnología no solo promueva informaciones verídicas, sino que también respete los derechos fundamentales y se alinee con los principios democráticos.

Además, es imperativo desarrollar e implementar políticas que aseguren una gobernanza efectiva de la IA, que reconozca y mitigue los riesgos mientras

maximiza sus beneficios. Esto incluye una colaboración internacional con amplia representación geográfica y la necesidad de una regulación que anticipe y responda a las rápidas evoluciones tecnológicas. Tales políticas deben garantizar que la IA sea una fuerza positiva para el fortalecimiento de las estructuras democráticas y no una amenaza.

Recientemente, la Asamblea General de las Naciones Unidas enfatizó la importancia de la cooperación internacional en la regulación de la IA, destacando la necesidad de integrar a los países en desarrollo en el proceso de formulación de políticas. Este enfoque global se hace relevante en virtud de que la desinformación y la manipulación tecnológica no respetan fronteras geográficas. La participación de los países en desarrollo garantiza que la gobernanza global de la IA refleje una gama más amplia de perspectivas y desafíos, contribuyendo a la creación de una sociedad mundial más inclusiva y equitativa.

Además, la resolución de la ONU subraya la necesidad de alinear el desarrollo y aplicación de la IA con los objetivos de desarrollo sostenible, insistiendo en que la tecnología debe promover un consenso sobre cuestiones esenciales de la sociedad. Esto resalta la importancia de una sociedad donde la tecnología amplíe las capacidades humanas y promueva la inclusión social, permitiendo un debate público más informado y equitativo.

En conclusión, la inteligencia artificial tiene el potencial de reforzar las democracias, siempre que su desarrollo y aplicación sean guiados por una estructura ética y regulatoria sólida que promueva la equidad, la justicia y la transparencia. Las políticas públicas y las iniciativas regulatorias globales deben reflejar estos valores, garantizando que la tecnología sirva a los ideales democráticos y contribuya a una gobernanza global más justa y eficaz, particularmente a través de la inclusión de países en desarrollo en el debate regulatorio.

Referencias

- Adams, Z., Osman, M., Bechlivanidis, C., & Meder, B. (2023). (Why) is misinformation a problem? *Perspectives on Psychological Science*, 18(6), 1436-1463. <https://doi.org/10.1177/17456916221141344>
- Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press.
- Cardenuto, J. P., Yang, J., Padilha, R., Wan, R., Moreira, D., Li, H., Wang, S., Andaló, F., Marcel, S., & Rocha, A. (2023). The age of synthetic realities: Challenges and opportunities. *APSIPA Transactions on Signal and Information Processing*, 12, e44. <https://doi.org/10.1561/116.00000138>
- Castells, M. (2006). *A Sociedade em Rede* (9.ª ed. rev. ampl.). Paz e Terra.
- Castro, L. M. C. de, & Cordeiro, M. E. B. (2022). Democracia e algoritmos: o papel das plataformas de mídias sociais na deterioração do Estado democrático brasileiro. En A. M. Dutra Ferreira da Silva, B. A. de Medeiros, D. Araújo, J. S. da Costa Neto, R. M. da Silva Melo, & R. Cavalcanti (orgs.), *Direito aplicado: tendências atuais do direito – Volume II* (pp. 213-231). Polimatia.
- Chen, S., Xiao, L., & Kumar, A. (2023). Spread of misinformation on social media: What contributes to it and how to combat it. *Computers in Human Behavior*, 141, 107643. <https://doi.org/10.1016/j.chb.2022.107643>
- Comisión Europea. (2021). *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial)*. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_1&format=PDF
- Da Silveira, S. A. (2019). *Democracia e os códigos invisíveis: Como os algoritmos estão modulando comportamentos e escolhas políticas*. Edições Sesc.
- European Commission. (2019). *Ethics guidelines for trustworthy AI*. <https://digital-strategy.ec.europa.eu/en/policies/ethics-guidelines-trustworthy-ai>
- European Commission. (2024). *Strengthened Code of Practice on Disinformation*. <https://digital-strategy.ec.europa.eu/policies/code-practice-disinformation>
- European Parliamentary Research Service. (2024). *Artificial Intelligence Act* (PE 698.792). [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)
- GPAI. (s. f.). *The Global Partnership on Artificial Intelligence*. <https://gpai.ai/>
- Habermas, J. (1989). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (T. Burger & F. Lawrence, trans.). The MIT Press.

- Instituto Igarapé. (2024, 9 de abril). *Instituto Igarapé e New America lançam guia sobre inteligência artificial segura para tomadores de decisão no Sul Global*. <https://igarape.org.br/instituto-igarape-e-new-america-lancam-guia-sobre-inteligencia-artificial-segura-para-tomadores-de-decisao-no-sul-global/>
- Kondamudi, M. R., Sahoo, S. R., & Yadav, N. (2023). Media-rich fake news detection: A survey. *Journal of King Saud University – Computer and Information Sciences*, 35, 101571. <https://doi.org/10.1016/j.jksuci.2023.101571>
- Mubarak, R., Als boui, T., Alshaikh, O., Inuwa-Dutse, I., & Bagheri, E. (2023). Leveraging AI to combat fake news: Techniques and challenges. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2023.3344653>
- Nan, X., Thier, K., & Wang, Y. (2023). Health misinformation: What it is, why people believe it, how to counter it. *Annals of the International Communication Association*, 47(4), 381-410. <https://doi.org/10.1080/23808985.2023.2225489>
- Neaher, G., Laforge, G., Muggah, R., & Seiler, G. (2024, abril). *Responsible and Safe AI: A Primer for Policymakers in the Global South*. Igarapé Institute. www.igarape.org.br
- OCDE. (2019). *Recommendation of the Council on Artificial Intelligence* [OCDE/LEGAL/0449]. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- OCDE. (2021). *Tools for Trustworthy AI: A Framework to Compare Implementation Tools* (OECD Digital Economy Papers, n.º 312).
- OCDE. (s. f.). *The OECD Artificial Intelligence Policy Observatory*. <https://oecd.ai/en/>
- Tufchi, S., Yadav, A., & Ahmed, T. (2023). A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities. *International Journal of Multimedia Information Retrieval*, 12(2), 28. <https://doi.org/10.1007/s13735-023-00296-3>
- Unión Europea. (2023). *Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo sobre normas armonizadas para un acceso justo a los datos y su utilización*. Diario Oficial de la Unión Europea. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202302854
- United Nations. (s. f.). *High-level Panel on Digital Cooperation*. <https://www.un.org/en/digital-cooperation-panel/>
- United Nations General Assembly. (2024). *Draft Resolution: Leveraging Safe, Secure, and Trustworthy Artificial Intelligence Systems for Sustainable Development (A/78/L.49)*. <https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf>
- Vasist, P. N., Chatterjee, D., & Krishnan, S. (2023). The polarizing impact of political disinformation and hate speech: A cross-country configural narrative. *Information Systems Frontiers*, 26(3), 663-688. <https://doi.org/10.1007/s10796-023-10390-w>
- World Economic Forum. (2024). *The Global Risks Report 2024*. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf