

**LA INCIDENCIA DE LA INTELIGENCIA ARTIFICIAL GENERATIVA EN LOS
DEEPFAKES Y SUS REPERCUSIONES EN LA CONFIANZA SOCIAL**
**THE IMPACT OF GENERATIVE ARTIFICIAL INTELLIGENCE ON DEEPFAKES AND
ITS REPERCUSSIONS ON SOCIAL TRUST**

Autores: ¹Monica Isabel Villamar Suastegui, ²Leopoldo Andrés Vera Pico, ³Marlene Guadalupe Castillo Pinargote y ⁴David Leonardo Reyes Romero, ⁵Oscar Alexander López Gorozabel.

¹ORCID ID: <https://orcid.org/0009-0000-2266-7509>

²ORCID ID: <https://orcid.org/0000-0002-8788-005X>

³ORCID ID: <https://orcid.org/0000-0002-7001-4382>

⁴ORCID ID: <https://orcid.org/0009-0005-2593-5966>

⁵ORCID ID: <https://orcid.org/0000-0002-0640-9953>

¹E-mail de contacto: monivisu@hotmail.com

²E-mail de contacto: leopoldo.vera@utm.edu.ec

³E-mail de contacto: marlene.castillo@utm.edu.ec

⁴E-mail de contacto: reyesromero887@gmail.com

⁵E-mail de contacto: oscar.lopez@utm.edu.ec

Afiliación: ^{1*4*}Investigador independiente, (Ecuador). ^{2*3*5*}Universidad Técnica de Manabí, (Ecuador).

Artículo recibido: 15 de Enero del 2026

Artículo revisado: 17 de Enero del 2026

Artículo aprobado: 30 de Enero del 2026

¹Ingeniero en Comercio Exterior. Actualmente cursa el Masterado en Administración de Empresas con Mención en Innovación y Emprendimiento de la Universidad Laica Eloy Alfaro de Manabí, (Ecuador).

²Ingeniero en Sistemas Informáticos. Máster de Investigación de Tecnologías de la Información mención en Seguridad de Redes y Comunicación.

³Ingeniera en Sistemas Informáticos. Máster de Investigación de Tecnologías de la Información mención en Seguridad de Redes y Comunicación.

⁴Ingeniero Eléctrico en Sistemas de Potencia. Máster en Energías Renovables. Actualmente cursa el Máster en Administración de Empresas con mención en Innovación y Emprendimiento.

⁵Ingeniero en Sistemas Informáticos. Licenciado en Trabajo Social. Máster en Ingeniería de Software. Máster en Intervención Social. Docente en la Universidad Técnica de Manabí, (Ecuador). Investigador.

Resumen

Esta investigación analiza el impacto disruptivo de la Inteligencia Artificial Generativa (IAG) en el ecosistema de los deepfakes y sus consecuencias para la confianza social. El estudio postula que la IAG actúa como un catalizador crítico, al democratizar y perfeccionar la creación de contenidos sintéticos hiperrealistas, lo que amplifica exponencialmente su potencial dañino. Metodológicamente, se adopta un enfoque cualitativo de tipo analítico-descriptivo, apoyado en la revisión sistemática de literatura académica, el análisis de casos de estudio emblemáticos y el examen de informes técnicos de empresas de ciberseguridad. Los hallazgos preliminares indican que la IAG ha eliminado barreras técnicas, permitiendo la generación masiva de deepfakes de alta calidad

para su uso en desinformación, suplantación de identidad y manipulación política. La discusión central gira en torno a la erosión de la confianza como bien público esencial: se identifica un debilitamiento de la confianza en los medios de comunicación, las instituciones políticas y las interacciones digitales básicas, fomentando un estado de escepticismo generalizado o realidad debilitada. Se concluye que la sociedad enfrenta una paradoja tecnológica donde la misma herramienta que impulsa la creatividad también amenaza los fundamentos de la verdad factual. La investigación aboga por la urgente necesidad de un marco multifacético que combine alfabetización digital crítica, desarrollo de herramientas de detección y marcos regulatorios ágiles para mitigar estas repercusiones y proteger el capital social en la era digital.

Palabras clave: Inteligencia artificial generativa, Deepfakes, Confianza social, Desinformación, Ética digital.

Abstract

This research analyzes the disruptive impact of Generative Artificial Intelligence (GAI) on the deepfake ecosystem and its consequences for social trust. The study posits that GAI acts as a critical catalyst, democratizing and refining the creation of hyperrealistic synthetic content, which exponentially amplifies its harmful potential. Methodologically, a qualitative, analytical-descriptive approach is adopted, supported by a systematic review of academic literature, the analysis of emblematic case studies, and the examination of technical reports from cybersecurity companies. Preliminary findings indicate that GAI has removed technical barriers, enabling the mass generation of high-quality deepfakes for use in disinformation, identity theft, and political manipulation. The central discussion revolves around the erosion of trust as an essential public good: a weakening of trust in the media, political institutions, and basic digital interactions is identified, fostering a state of widespread skepticism or a weakened sense of reality. The study concludes that society faces a technological paradox where the very tool that fosters creativity also threatens the foundations of factual truth. The research argues for the urgent need for a multifaceted framework that combines critical digital literacy, the development of detection tools, and agile regulatory frameworks to mitigate these repercussions and protect social capital in the digital age.

Keywords: Generative artificial intelligence, Deepfakes, Social trust, Disinformation, Digital ethics.

Sumário

Esta pesquisa analisa o impacto disruptivo da Inteligência Artificial Generativa (IAG) no ecossistema de deepfakes e suas consequências para a confiança social. O estudo postula que a IAG atua como um catalisador crucial, democratizando e refinando a criação de

conteúdo sintético hiper-realista, o que amplifica exponencialmente seu potencial nocivo. Metodologicamente, adota-se uma abordagem qualitativa, analítico-descritiva, apoiada por uma revisão sistemática da literatura acadêmica, a análise de estudos de caso emblemáticos e o exame de relatórios técnicos de empresas de cibersegurança. Resultados preliminares indicam que a IAG removeu barreiras técnicas, possibilitando a geração em massa de deepfakes de alta qualidade para uso em desinformação, roubo de identidade e manipulação política. A discussão central gira em torno da erosão da confiança como um bem público essencial: identifica-se um enfraquecimento da confiança na mídia, nas instituições políticas e nas interações digitais básicas, fomentando um estado de ceticismo generalizado ou um senso de realidade fragilizado. O estudo conclui que a sociedade enfrenta um paradoxo tecnológico em que a própria ferramenta que fomenta a criatividade também ameaça os fundamentos da verdade factual. A pesquisa defende a necessidade urgente de uma estrutura multifacetada que combine alfabetização digital crítica, o desenvolvimento de ferramentas de detecção e marcos regulatórios ágeis para mitigar essas repercuções e proteger o capital social na era digital.

Palavras-chave: Inteligencia artificial generativa, Deepfakes, Confiança social, Desinformação, Ética digital.

Introducción

Vivimos en una era de transformación digital sin precedentes, donde la Inteligencia Artificial (IA) ha dejado de ser un concepto de la ciencia ficción para convertirse en un pilar fundamental del progreso tecnológico. Entre sus avances más disruptivos se encuentra la Inteligencia Artificial Generativa (en adelante IAG), sistemas capaces de crear contenido nuevo y original como texto, imágenes, audio y video a partir de los patrones que aprenden de vastos conjuntos de datos. Esta capacidad, que promete revolucionar campos como el arte, el

diseño, la medicina y la educación, representa simultáneamente un arma de doble filo de enorme potencia. El lado más oscuro y preocupante de esta innovación se manifiesta en la evolución exponencial de los deepfakes. Estas técnicas de suplantación hiperrealista, que antes requerían importantes recursos técnicos y tiempo, se han democratizado y perfeccionado dramáticamente gracias a la IAG. Plataformas de acceso público permiten ahora, con apenas unos clics y una simple fotografía, generar videos falsos, pero visualmente convincentes y audios sintéticos prácticamente indistinguibles de la realidad. Esta facilidad de creación ha catapultado la calidad, la cantidad y la velocidad de producción de contenidos manipulados, erosionando la ya frágil frontera entre lo verdadero y lo falso.

Esta distorsión de la realidad digital plantea una amenaza profunda para los cimientos de la confianza social. La confianza, entendida como el adhesivo que mantiene unidas las relaciones interpersonales, las instituciones y los sistemas de información, se ve severamente comprometida. Cuando la ciudadanía no puede creer en lo que ve o escucha, se socava la credibilidad de figuras públicas, se debilita el periodismo, se manipulan procesos electorales y se vulnera la dignidad de las personas mediante suplantaciones maliciosas. El panorama resultante es el de una esfera pública intoxicada por la desinformación, donde la duda y la incredulidad se generalizan, dando paso a lo que algunos expertos han denominado una "crisis de la realidad" o un "nihilismo epistemológico". Por lo tanto, esta investigación se propone analizar de manera crítica y exhaustiva la incidencia directa de la Inteligencia Artificial Generativa en el desarrollo y proliferación de deepfakes, y evaluar sus repercusiones concretas en los niveles de confianza social. A través de un

examen del estado del arte de la tecnología, el análisis de casos emblemáticos y la revisión de estudios sobre el impacto psicosocial de la desinformación, este trabajo busca contribuir a la comprensión de uno de los desafíos más urgentes de nuestro tiempo, además de sentar las bases para discutir posibles marcos de gobernanza, alfabetización digital y soluciones tecnológicas que permitan mitigar estos efectos y proteger la integridad de nuestro tejido social en la era digital.

La IAG se consolidó en la última década como una de las áreas más innovadoras de la inteligencia artificial, permitiendo la creación de textos, imágenes, audios y videos sintéticos de alta calidad mediante arquitecturas como Generative Adversarial Networks (en adelante GANs) y modelos de difusión (Chesney y Citron, 2019). Entre sus aplicaciones más controvertidas se encuentran los deepfakes, entendidos como contenidos audiovisuales manipulados mediante algoritmos de aprendizaje profundo que simulan escenarios inexistentes con un alto grado de realismo (Vaccari y Chadwick, 2020). La emergencia de estas tecnologías plantea serias preocupaciones respecto a la erosión de la confianza social y la propagación de desinformación, al alterar los mecanismos tradicionales de verificación de información en ámbitos políticos, sociales y económicos (Natale, 2024). La revisión sistemática de literatura en Scopus, Web of Science y JSTOR, usando palabras clave como "generative AI", "deepfakes", "social trust" y "disinformation", permitió identificar tres dimensiones centrales en el debate académico: incidencia social de deepfakes, informes técnicos y white papers, y casos de estudio en diferentes ámbitos sociales.

Diversos estudios señalan que la exposición a deepfakes disminuye la confianza en

instituciones políticas y medios de comunicación, incluso cuando los individuos reconocen la manipulación (Hoi y Quang, 2025). Se observa también el fenómeno conocido como “efecto de mentira plausible”, donde la simple posibilidad de existencia de deepfakes debilita la credibilidad del discurso público y facilita la propagación de desinformación (Marwick y Lewis, 2023). Investigaciones recientes confirman que la confianza en el gobierno y la familiaridad con tecnologías digitales pueden moderar la percepción de riesgo ante deepfakes, aunque no eliminan la susceptibilidad a la desinformación (Hynek et al., 2025; Jin et al, 2023). Asimismo, el carácter viral de estos contenidos en redes sociales dificulta su detección y corrección, amplificando su impacto social (Drolsbach y Pröllochs, 2025; Adel, 2024). La exposición repetida a deepfakes incrementa la normalización de la desinformación, generando un entorno social donde la veracidad se percibe como relativa, afectando la confianza generalizada en la información digital (Yam y Skorburg, 2021; Murguía et al. 2025).

Diversos informes técnicos destacan los riesgos concretos de los deepfakes. Europol (2022) advirtió sobre su uso en fraudes financieros, manipulación electoral y delitos sexuales digitales, mientras que Sensity (2023) evidenció que la mayoría de los deepfakes detectados en la red correspondían a pornografía no consensuada, revelando un patrón de victimización con fuerte sesgo de género. Empresas tecnológicas y consorcios, como Microsoft y Partnership on AI (2023), enfatizaron la necesidad de marcos de trazabilidad y etiquetado digital, mientras que think tanks como Brookings y Knight Foundation (2022) sugirieron políticas públicas que integren alfabetización mediática, regulaciones claras y cooperación internacional.

Los informes recientes de la ITU (2025) reiteran la urgencia de estándares globales de autenticación multimedia y herramientas de verificación digital para preservar la confianza pública (Villasenor, 2019; Lundberg y Mozelius, 2025). Los hallazgos teóricos se ejemplificaron mediante tres casos emblemáticos que evidenciaron la transversalidad del fenómeno de los deepfakes en distintos ámbitos sociales. En el terreno social, la proliferación de contenidos pornográficos sintéticos no consensuales representó una de las principales aplicaciones maliciosas de esta tecnología, afectando a miles de víctimas y visibilizando la ausencia de marcos legales eficaces para su protección, lo que ha sido documentado en estudios y reportes recientes (Sensity, 2023; Europol, 2022).

En el ámbito político, el video manipulado que mostraba al presidente francés Emmanuel Macron en un metraje alterado en 2024 ilustró el potencial de los deepfakes para viralizarse en redes sociales y erosionar la confianza ciudadana en procesos democráticos, fenómeno ya advertido por investigaciones sobre desinformación digital y confianza pública (Fallis, 2021). Finalmente, en el plano fraudulento, el caso registrado en 2019 en el Reino Unido, en el que delincuentes clonaron la voz del director ejecutivo de una empresa energética para ordenar una transferencia de 243.000 dólares, constituyó uno de los primeros fraudes financieros confirmados mediante el uso de audio sintético, evidenciando la vulnerabilidad de los sistemas de verificación tradicionales frente a este tipo de ataques (Mounica, 2025; Somorjeet y Lungharvanao, 2025). Estos ejemplos reflejan la necesidad de marcos regulatorios y técnicos sólidos, alfabetización mediática y cooperación internacional para mitigar los riesgos asociados a la IAG y los deepfakes, garantizando la

integridad de la información y la confianza social (Brown, 2020; Al-Khazraji et al., 2023).

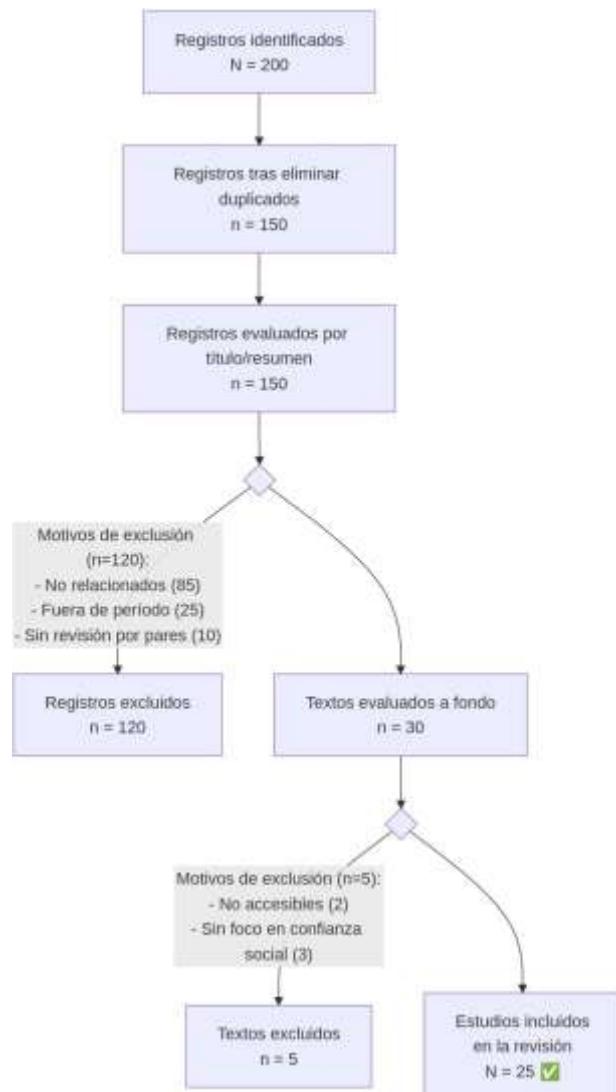
Luego de la revisión de literatura académica y los informes técnicos, se coincide en que los deepfakes representan una amenaza multifacética para la sociedad contemporánea. No solo facilitan la circulación de desinformación y la manipulación en contextos de alta sensibilidad política y social, sino que también generan daños concretos a nivel individual a través del fraude y la vulneración de la intimidad. Frente a este escenario, se plantea la necesidad de un abordaje integral que articule el desarrollo de tecnologías de detección, la formulación de marcos regulatorios adaptativos, la promoción de una ética digital robusta y la implementación de programas de alfabetización mediática. De esta manera, se reconoce que el principal desafío no residía únicamente en la sofisticación técnica que hacía cada vez más verosímiles a los deepfakes, sino en su capacidad de minar la confianza social y multiplicar los vectores de desinformación en un entorno digital globalizado e hiperconectado.

Materiales y Métodos

Para abordar el complejo fenómeno de estudio, esta investigación se diseñó bajo un enfoque cualitativo de tipo analítico-descriptivo, ideal para comprender las relaciones profundas entre la tecnología, su aplicación y su impacto social. El diseño se implementó como un estudio de caso múltiple y análisis documental, que permitió una exploración exhaustiva y multidimensional. La recolección de datos se realizó mediante una revisión sistemática de literatura en bases de datos académicas especializadas (Scopus, Web of Science, JSTOR), utilizando un protocolo definido con palabras clave como "generative AI", "deepfakes", "social trust" y "disinformation".

Esta búsqueda se complementó con el análisis de informes técnicos y white papers de organizaciones líderes en ciberseguridad e inteligencia artificial, lo que proporcionó una comprensión actualizada de las capacidades técnicas y las tendencias. Adicionalmente, se seleccionaron y analizaron en profundidad tres casos emblemáticos de deepfakes en los ámbitos político, social y fraudulento, los cuales sirvieron como evidencia concreta para ilustrar los patrones de creación, difusión e impacto.

Figura 1. Diagrama de flujo



Los datos cualitativos obtenidos de todas estas fuentes fueron procesados mediante análisis de contenido temático, codificando la información

para identificar categorías, temas y relaciones recurrentes, lo que aseguró un análisis riguroso y estructurado de la información.

Resultados y Discusión

El análisis de los datos arrojó resultados claros y contundentes. En primer lugar, se confirmó que la IAG actúa como un catalizador decisivo, habiendo democratizado y perfeccionado la creación de deepfakes al eliminar las barreras técnicas y de recursos que antes existían. Los casos analizados demostraron una clasificación tangible del impacto en tres ámbitos principales: en el político, se documentaron campañas de desinformación diseñadas para desacreditar oponentes y manipular la opinión pública; en el social y personal, se evidenció el uso malicioso para fraudes por suplantación de identidad y acoso mediante pornografía no consensuada, con graves daños a la reputación y salud mental de las víctimas; y en el periodístico, se verificó la presencia del "efecto liar's dividend" (el dividendo del mentiroso), donde un video generado por IA puede ser utilizado por actores de mala fe para desmentir evidencias reales e intactas. Finalmente, se mapeó con claridad cómo estos deepfakes de alta calidad se integran y son amplificados por ecosistemas de desinformación preexistentes, utilizando redes de bots y comunidades online para alcanzar una viralización rápida y un impacto social máximo.

Se argumenta que el fenómeno deepfake trasciende lo técnico para instalar una crisis epistemológica, donde la erosión de la capacidad para discernir lo verdadero de lo falso socava la realidad compartida, base del contrato social, fomentando un escepticismo generalizado que paraliza el discurso público. Autores como Palomo (2021) han señalado la “confusión epistemológica” generada por la desinformación digital, donde no se da valor a

la evidencia objetiva sino a la creencia. Asimismo, estudios como Generative AI and deepfakes: a human rights approach to tackling harmful content (Romero, 2024) destacan cómo regulaciones tardías ante tecnologías generativas atentan contra los derechos fundamentales. Esta investigación sitúa en el centro del debate la paradoja de la democratización tecnológica: la misma herramienta que impulsa la creatividad y la innovación es también un instrumento de desestabilización social, planteando una tensión inherente entre la libertad de expresión, la seguridad nacional y los derechos individuales. En Combatting deepfakes: Policies to address national security threats and rights violations (Miotti y Wasil, 2024) se resaltan los riesgos que los deepfakes introducen para la seguridad nacional y la libertad individual si no se actúa integralmente.

Asimismo, se discuten las limitaciones de las soluciones unidimensionales, subrayando que la carrera armamentística entre la creación y la detección de deepfakes es insuficiente por sí sola. Estudios técnicos recientes como Deepfake Forensic Analysis: Source Dataset Attribution and Legal Implications of Synthetic Media Manipulation (Cassia et al., 2025) muestran que aunque hay avances muy altos en detección, la atribución de origen y la regulabilidad legal siguen siendo débiles si no se combinan con políticas sociales. Frente a este panorama, la investigación concluye que la respuesta debe ser necesariamente multifacética y orientada a construir resiliencia social. Se propone un marco de acción integrado por tres pilares:

- La implementación urgente de programas robustos de alfabetización mediática y pensamiento crítico dirigidos a la ciudadanía; ideas

presentes en la literatura de regulación y ética de (Demkin, 2024).

- La inversión y desarrollo colaborativo de herramientas de verificación accesibles para periodistas, fact-checkers y plataformas digitales; como sugieren Miotti y Wasil (2024) y Cassia et al. (2025).
- El diseño de marcos regulatorios ágiles y proporcionales que persigan a los actores malintencionados y establezcan responsabilidades claras, sin coartar la innovación legítima (Ginsburg y Austin, 2025), que comparan jurisdicciones y modelos jurídicos para equilibrar derechos y seguridad.

Conclusiones

La IAG ha operado como un catalizador disruptivo, transformando radicalmente el panorama de la desinformación digital. Al democratizar el acceso a la creación de contenidos sintéticos hiperrealistas, ha eliminado las barreras técnicas y de recursos que anteriormente limitaban la producción de deepfakes, facilitando su generación masiva, acelerando su ritmo de producción y elevando su calidad hasta niveles de verosimilitud alarmantes. Esto ha amplificado exponencialmente su potencial de daño al escalar su uso malintencionado. En segundo término, el análisis permitió categorizar y verificar el impacto multifacético de esta tecnología. Los resultados demuestran de manera contundente que los deepfakes potenciados por IAG no son una amenaza abstracta, sino una realidad operativa en ámbitos concretos: su utilización en la manipulación política erosiona los procesos democráticos; su aplicación en el fraude y el acoso causa daños personales y patrimoniales

graves; y su existencia debilita la labor periodística a través del "efecto liar's dividend", socavando la credibilidad de cualquier evidencia audiovisual.

Además, se ha comprobado que el efecto más corrosivo y trascendental recae sobre los cimientos de la confianza social. La IAG, al borrar progresivamente la frontera entre lo real y lo sintético, impulsa una crisis epistemológica donde la duda se generaliza. Esto da lugar a un estado de realidad debilitada que fractura el consensus reality necesario para el funcionamiento de la sociedad, debilitando la confianza interpersonal, en las instituciones y en los sistemas de información. Finalmente, se concluye que enfrentar este desafío requiere superar el enfoque puramente tecnológico de la detección. La solución yace en la construcción de una resiliencia social digital multifacética. Por lo tanto, esta investigación aboga por la implementación urgente de un marco integrado que combine: 1) la alfabetización digital crítica para empoderar a la ciudadanía, 2) el desarrollo colaborativo de herramientas de verificación y 3) marcos regulatorios ágiles que disuadan el uso malintencionado protegiendo los derechos fundamentales. El futuro de la confianza en la esfera digital depende de nuestra capacidad colectiva para responder a esta paradoja tecnológica con sabiduría, equilibrio y una acción coordinada.

Referencias Bibliográficas

- Adel, M. (2024). *Deepfakes and media integrity: Navigating the new reality of synthetic content*. *Journal of Media and Interdisciplinary Studies*, 9(2), 48–94.
https://jmis.journals.ekb.eg/article_380204_8179b07978fe29bc28e9aa33600a5d9c.pdf
- Al-Khazraji, S., Saleh, H., & Mishkhal, I. (2023). *Impact of deepfake technology on social media: Detection, misinformation and societal implications*. *The Eurasia*

- Proceedings of Science, Technology, Engineering & Mathematics*, 23(1), 429–441.
<https://dergipark.org.tr/en/download/article-file/3456697>
- Brookings Institution. (2022). Deepfakes and democracy: Policy recommendations. Brookings Institution.
<https://www.brookings.edu/articles/deepfakes-and-international-conflict/>
- Brown, N. (2020). *Deepfakes and the weaponization of information*. *Virginia Journal of Law & Technology*, 23(1), 1–57.
https://static1.squarespace.com/static/5e793709295d7b60295b2d29/t/5f0b7e4e0f4f095f7e70613f/1594588750542/v23i1_1-Brown.pdf
- Cassia, M., Guarnera, L., Casu, M., Zangara, I., & Battiatto, S. (2025). *Deepfake forensic analysis: Source dataset attribution and legal implications of synthetic media manipulation*. arXiv.
<https://arxiv.org/abs/2505.11110>
- Chesney, R., & Citron, D. (2019). *Deepfakes and the new disinformation war: The coming age of post-truth geopolitics*. *Foreign Affairs*, 98(1), 1–10.
<https://www.jstor.org/stable/26798018>
- Citron, D., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. University of California Berkeley School of Law.
https://scholarship.law.bu.edu/faculty_scholarship/640
- Demkin, V. (2024). *Deepfakes: Search for a model of legal regulation*. *Legal Issues in the Digital Age*, 5(4), 73–91.
<https://doi.org/10.17323/2713-2749.2024.4.73.91>
- Drolsbach, S., & Pröllochs, N. (2025). *Characterizing AI-generated misinformation on social media*. *Social and Information Networks*, 7(1), 1–12.
<https://arxiv.org/html/2505.10266v1>
- Europol. (2022). Deepfakes: A new threat to public safety. European Union Agency for Law Enforcement Cooperation.
https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf
- Fallis, D. (2021). *The epistemic threat of deepfakes*. *Philosophy & Technology*, 34(4), 623–643.
<https://doi.org/10.1007/s13347-020-00419-2>
- Ginsburg, J., & Austin, G. (2025). *Deepfakes in domestic and international perspective*. *Columbia Journal of Law & the Arts*, 48(3), 297–306.
<https://doi.org/10.52214/jla.v48i3.13868>
- Hoi, N., & Quang, B. (2025). *The impact of deepfake technology on public trust in retail systems*. *International Journal of Education and Social Science Research*.
<https://doi.org/10.37500/IJESSR.2025.8321>
- Hynek, N., Gavurova, B., & Kubak, M. (2025). *Risks and benefits of artificial intelligence deepfakes: Systematic review and comparison of public attitudes in seven European countries*. *Journal of Innovation & Knowledge*, 10(5), 1–20.
<https://doi.org/10.1016/j.jik.2025.100782>
- International Telecommunication Union. (2025). Global standards for digital authentication. ITU.
<https://www.itu.int/hub/publication/t-ai4g-ai4good-2024-7/>
- Jin, X., Zhang, Z., Gao, B., Gao, S., Zhou, W., Yu, N., & Wang, G. (2023). *Assessing the perceived credibility of deepfakes: The impact of system-generated cues and video characteristics*. *New Media & Society*, 27(3), 1651–1672.
<https://doi.org/10.1177/14614448231199664>
- Knight Foundation. (2022). Protecting truth: Addressing the deepfake challenge.
<https://knightfoundation.org/wp-content/uploads/2022/03/KMAD-2022-1.pdf>
- Lundberg, E., & Mozelius, P. (2025). *The potential effects of deepfakes on news media and entertainment*. *AI & Society*, 40, 2159–2170.
<https://doi.org/10.1007/s00146-024-02072-1>

- Marwick, A., & Lewis, R. (2023). Media manipulation and disinformation online. Data & Society Research Institute. <https://datasociety.net/library/media-manipulation-and-disinfo-online/>
- Miotti, A., & Wasil, A. (2024). *Combatting deepfakes: Policies to address national security threats and rights violations*. arXiv. <https://doi.org/10.48550/arXiv.2402.09581>
- Murguía, J., Toba, C., Campos, S., Aramburo, L., & Diaz, G. (2025). *Deepfakes: Revisión sistemática de tecnologías, impacto y estrategias de detección*. Revista de Investigación en Tecnologías de la Información, 13(29), 19–37. <https://doi.org/10.36825/RITI.13.29.003>
- Natale, S. (2024). *Digital media and the banalization of deception*. Convergence, 31(1), 402–419. <https://doi.org/10.1177/1354856524131178>
- Palomo, M. (2021). *Incidencias filosóficas actuales en la sociedad digital: Ideologías, desinformación y confusión epistemológica*. Arbor, 197(802), a630. <https://doi.org/10.3989/arbor.2021.802008>
- Romero, F. (2024). *Generative AI and deepfakes: A human rights approach to tackling harmful content*. International Review of Law, Computers and Technology, 38(3), 297–326. <https://doi.org/10.1080/13600869.2024.2324540>
- Vaccari, C., & Chadwick, A. (2020). *Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news*. Social Media + Society, 6(1). <https://doi.org/10.1177/2056305120903408>
- Villasenor, J. (2019). Artificial intelligence, deepfakes, and the uncertain future of truth. Brookings Institution Report. <https://www.brookings.edu/articles/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/>
- Yam, J., & Skorburg, J. (2021). *From human resources to human rights: Impact assessments for hiring algorithms*. Ethics and Information Technology, 23, 611–623. <https://doi.org/10.1007/s10676-021-09599-7>



Esta obra está bajo una licencia de Creative Commons Reconocimiento-No Comercial 4.0 Internacional. Copyright © Monica Isabel Villamar Suastegui, Leopoldo Andrés Vera Pico, Marlene Guadalupe Castillo Pinargote y David Leonardo Reyes Romero, Oscar Alexander López Gorozabel.

