https://doi.org/10.69639/arandu.v12i3.1443

Impacto de los *deepfakes* en la confianza pública y la credibilidad de los medios de comunicación una revisión sistemática

Impact of deepfakes on public trust and media credibility: a systematic review

Noemi Katiuska Salinas Parra
nsalinasp@unemi.edu.ec
https://orcid.org/0009-0005-0693-4175
Universidad Estatal de Milagro
Milagro, Ecuador

Ángela Victoria Salcedo Pinela
asalcedop2@unemi.edu.ec
https://orcid.org/0000-0001-8519-768X
Universidad Estatal de Milagro
Milagro, Ecuador

Artículo recibido: 18 julio 2025 - Aceptado para publicación: 28 agosto 2025
Conflictos de intereses: Ninguno que declarar.

RESUMEN

Los deepfakes, generados mediante técnicas de aprendizaje profundo, han demostrado ser una herramienta poderosa para manipular información visual y auditiva, lo que plantea desafíos significativos para la verificación de la verdad y la estabilidad democrática. Nuestro análisis examina cómo los deepfakes afectan la percepción pública, erosionan la confianza en las instituciones y alteran los procesos electorales. Además, se discuten las implicaciones jurídicas y tecnológicas de esta tecnología, así como las estrategias necesarias para contrarrestar sus efectos negativos. La revisión sistemática de artículos académicos de revistas Q1, Q2 y Q3 revela que los deepfakes representan una amenaza creciente para la confianza pública y la credibilidad de los medios de comunicación. La capacidad de los deepfakes para manipular información ha llevado a una erosión de la confianza en las fuentes mediáticas establecidas y en las instituciones gubernamentales. Es crucial desarrollar tecnologías de detección más avanzadas y estrategias de regulación efectivas para contrarrestar los efectos negativos de los deepfakes. La educación pública sobre los riesgos asociados con esta tecnología es esencial para fortalecer la resistencia informativa y proteger la integridad de los procesos democráticos.

Palabras clave: deepfakes, desinformación, confianza pública, democracia

ABSTRACT

Deepfakes, generated using deep learning techniques, have proven to be a powerful tool for manipulating visual and auditory information, posing significant challenges for truth verification

and democratic stability. Our analysis examines how deepfakes affect public perception, erode trust in institutions, and disrupt electoral processes. In addition, the legal and technological implications of this technology are discussed, as well as the strategies needed to counteract its negative effects. A systematic review of academic articles from Q1, Q2, and Q3 journals reveals that deepfakes pose a growing threat to public trust and media credibility. The ability of deepfakes to manipulate information has led to an erosion of trust in established media sources and government institutions. It is crucial to develop more advanced detection technologies and effective regulatory strategies to counteract the negative effects of deepfakes. Public education about the risks associated with this technology is essential to strengthen information resilience and protect the integrity of democratic processes

Keywords: deepfakes, disinformation, public trust, democracy

INTRODUCCIÓN

La proliferación de contenidos audiovisuales manipulados digitalmente, conocidos como deepfakes o ultrafalsos, ha emergido como una de las amenazas más apremiantes y multifacéticas en la actual era digital (Barrientos-Báez et al., 2024). Estos contenidos, que simulan de manera hiperrealista la apariencia y el habla de personas reales (García-Ull, 2021a), son generados mediante avanzadas técnicas de Inteligencia Artificial (IA), predominantemente a través del aprendizaje profundo (deep learning) y las Redes Generativas Antagónicas (GANs) (Khan et al., 2025). El término "deepfake" se acuñó combinando "deep learning" y "fake", reflejando su naturaleza sintética y engañosa (Lendvai & Gosztonyi, 2024). Su aparición en internet en 2017 marcó un antes y un después, democratizando la manipulación digital y haciéndola accesible incluso a usuarios sin experiencia técnica (Barrientos-Báez et al., 2024)

El impacto de los deepfakes es profundo y perturbador, generando una erosión significativa de la confianza pública y la credibilidad en los medios de comunicación y las instituciones gubernamentales (Momeni, 2025). Esta tecnología dificulta cada vez más la distinción entre información verídica y manipulada, contribuyendo a una creciente desconfianza generalizada, incluso hacia fuentes reconocidas (Weikmann et al., 2025). La pérdida de credibilidad de los medios digitales se ve acelerada por la capacidad de los deepfakes de minar el pensamiento crítico lo cual se vuelve un mayor reto para cualquier medio incluso los medios tradicionales. Como señala (García-Ull, 2021b) la mayor amenaza no reside tanto en el engaño directo al receptor, sino en la pérdida total de credibilidad de la información misma. Ejemplos de deepfakes de personalidades como Barack Obama, Donald Trump y Mark Zuckerberg han acaparado la atención mediática internacional, ilustrando la seriedad de esta amenaza (García-Ull, 2021b). Esto evidencia el alto potencial de estas tecnologías para manipular la opinión pública y erosionar la confianza en los medios. Su realismo genera confusión entre lo verdadero y lo falso, poniendo en riesgo la credibilidad informativa y destacando la necesidad urgente de regulación, alfabetización mediática y desarrollo de tecnologías de detección para mitigar su impacto en la esfera pública y democrática. Más allá de la desinformación general, los deepfakes plantean serias amenazas para la privacidad y la seguridad personal, incrementando los riesgos derivados de la suplantación de identidad (Lendvai & Gosztonyi, 2024). Por lo cual esta tecnología permite crear contenido falso convincente que simula la apariencia y voz de individuos, utilizándose para generar pornografía no consentida, facilitar fraudes financieros y causar daño reputacional. Su creciente sofisticación hace difícil distinguirlos de la realidad, lo que erosiona la confianza pública y la seguridad personal. Un ámbito particularmente preocupante es la violencia política de género, donde los deepfakes se utilizan para atacar y desacreditar a las mujeres políticas, a menudo a través de la hipersexualización y la difusión de pornografía no consentida (Barrientos-Báez et al., 2024; Shibuya et al., 2025). Estas prácticas se inscriben en una lucha de poder que perpetúa la misoginia

y mantiene las estructuras de poder existentes lo cual es una forma de troleo de género para silenciar voces críticas y disuadir su participación pública, reforzando la mirada patriarcal.

La lucha contra los deepfakes se ve obstaculizada por la constante sofisticación de las técnicas de generación, lo que hace que los sistemas de detección actuales no siempre logren mantenerse al ritmo de su evolución (García-Ull, 2021a; Liu et al., 2024). Los desarrolladores de deepfakes utilizan los resultados de las investigaciones publicadas para mejorar su tecnología y evadir los nuevos sistemas de detección. Aunque se han identificado indicadores sutiles para detectar deepfakes, como imperfecciones en el brillo, distorsión facial, o anomalías en el parpadeo , la IA generativa puede producir contenido tan convincente que es difícil de discernir por el ojo humano (Tran et al., 2024). Esto subraya la necesidad crítica de desarrollar tecnologías de detección más avanzadas y robustas además el análisis multimodal, que combina datos textuales, visuales y auditivos, puede mejorar las capacidades de detección al identificar inconsistencias entre modalidades.

Los resultados de revisiones sistemáticas, como la realizada con una exhaustiva indagación de artículos científicos de bases de datos como Scopus, evidencian un interés académico creciente y global en la temática. Se observa un aumento significativo en las publicaciones sobre deepfakes y sus implicaciones a partir de 2024, lo que sugiere la consolidación de este campo de investigación como uno de alta relevancia, posiblemente impulsado por los avances tecnológicos, el aumento de casos de desinformación y la necesidad de marcos éticos y legales. Países como China y España lideran la producción científica en este ámbito, reflejando la necesidad compartida de comprender y afrontar las implicaciones sociopolíticas y comunicativas del fenómeno a nivel mundial. Esto incluye el estudio de las diferencias lingüísticas entre el contenido generado por humanos y el generado por IA, donde se ha demostrado que características sintácticas y léxicas son clave para la detección (Tran et al., 2024)

MATERIALES Y MÉTODOS

La investigación es cualitativa a través de una revisión sistemática de la literatura, se realizó indagación exhaustiva de artículos científicos, informes y documentos pertinentes se centró en el impacto de los deepfakes en la confianza pública y credibilidad de los medios de comunicación, tanto a nivel nacional como internacional., de base de datos Scopus como soporte científico y evidencia de calidad.

Se extrajo la información de aporte en la investigación de esta cadena de búsqueda:

TITLE-ABS-KEY ("deepfake*" OR "synthetic media" OR "AI-generated content" OR "fake media" OR "artificial intelligence AND misinformation") AND TITLE-ABS-KEY ("trust" OR "public confidence" OR "credibility" OR "perception") AND TITLE-ABS-KEY ("news" OR "journalism" OR "media" OR "social media" OR "mass communication") AND PUBYEAR > 1999 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (

LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "PSYC") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (EXACTKEYWORD , "Deepfake") OR LIMIT-TO (EXACTKEYWORD , "Disinformation") OR LIMIT-TO (EXACTKEYWORD , "Deepfakes") OR LIMIT-TO (EXACTKEYWORD , "Artificial Intelligence"))

Estrategia de búsqueda y fuentes de información

La fuente de información que se utilizó como base de datos es Scopus se inició la búsqueda general sobre **deepfakes, confianza pública y medios de comunicación**, en periodo de publicación esta es del 2000 en adelante, el tipo de documento solo se incluyen artículos de investigación ("ar") y revisiones sistemáticas ("re"), en el idioma se priorizan artículos en inglés y español, dado que son los más accesibles en literatura académica. La cual al inicio obtuvo 58 documentos sin ningún tipo de rango en años de publicación, en las sub áreas se filtró por ciencias sociales, informática, empresa, gestión y contabilidad, psicología y ciencias medioambientales además en tipo de documento se seleccionó artículos y revisión.

Selección de estudios y criterios de elegibilidad

A continuación, se empezó a filtrar los criterios para un resultado más detallado, tal como se muestra en la tabla 1.

Tabla 1

Muestra de artículos seleccionados

No	Criterios aplicados en la revisión sistemática
1	La fecha de publicación se establece desde el año 2020 a 2025 para la revisión.
2	El área de temática se limitó a informática y ciencias sociales
3	El tipo de documento a utilizar serán solo artículos
4	Las palabras clave se asocian a las variables del estudio.
5	El idioma del estudio corresponde a los admitidos para la revisión en el idioma inglés y español
6	La investigación tiene una versión completa (open access)
7	La investigación evidencia resultados o indicadores cualitativo

Una vez aplicados los criterios de búsqueda y los filtros antes mencionados la investigación se redujo a 28 documentos lo cual permitirá una mayor profundización del tema. Asimismo, se procedió a tabular y organizar el contenido de los estudios mediante el uso de la herramienta Zotero que permite tener un orden más claro. El objetivo de la investigación es determinar cómo los deepfakes erosionan la confianza del público en los medios de comunicación tanto en medios tradicionales como digitales y como afecta la percepción general sobre la credibilidad periodística.

RESULTADOS

Tabla 2

Muestra de artículos seleccionados

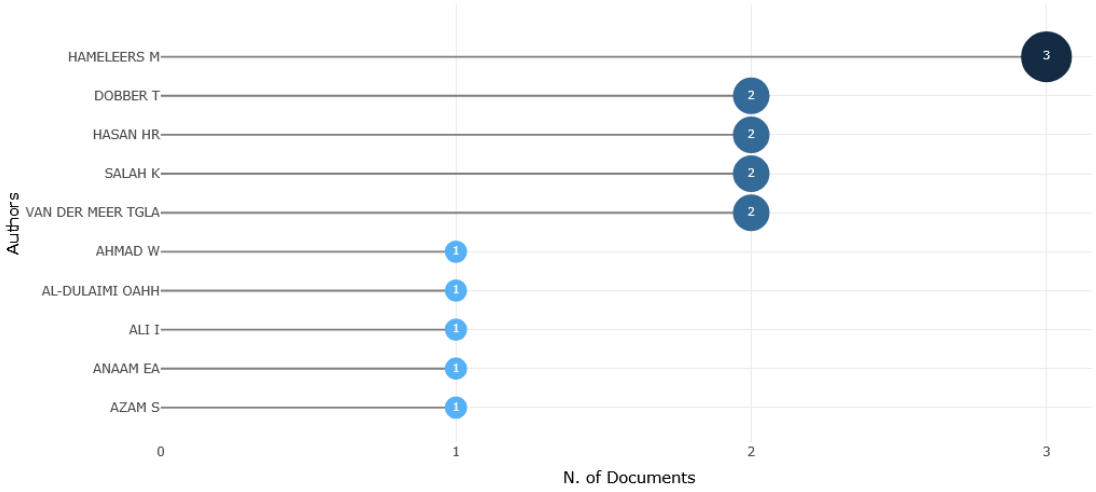
Nº	Autores	Título de la investigación
1.	Xu, Z., Wen, X., Zhong, G., Fang, Q.	Public perception towards deepfake through topic modelling and sentiment analysis of social media data
2	Shibuya, Y., Nakazato, T., Takagi, S.	How do people evaluate the accuracy of video posts when a warning indicates they were generated by AI?
3	Momeni, M.	Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions Through Misinformation Momeni, M.
4	Chen, G., Du, C., Yu, Y., ... Duan, H., Zhu, H.	A Deepfake Image Detection Method Based on a Multi-Graph Attention Network
5	Yee, L.F., Kumaresan, S.P., Anaam, E.A., Kong, W.C.	Social Platforms in the Deepfake Age: Navigating Media Trust through Media Literacy
6	Weikmann, T., Greber, H., Nikolaou, A.	After Deception: How Falling for a Deepfake Affects the Way We See, Hear, and Experience Media
7	Hameleers, M., van der Meer, T., Vliegenthart, R.	How persuasive are political cheapfakes disseminated via social media? The effects of out-of-context visual disinformation on message credibility and issue agreement
8	Ölvecký, M., Huraj, L., Brlej, I.	Evaluating the Effectiveness of Deepfake Video Detection Tools: A Comparative Study
9	Rubio, R.	Does artificial intelligence redefine election campaigns and their democratic effects? | EL USO DE LA INTELIGENCIA ARTIFICIAL EN LAS CAMPAÑAS ELECTORALES Y SUS EFECTOS DEMOCRÁTICOS
10	Tran, V.H., Sebastian, Y., Karim, A., Azam, S.	Distinguishing Human Journalists from Artificial Storytellers Through Stylistic Fingerprints
11	Sanchez-Acedo, A., Carbonell-Alcocer, A., Gertrudix, M., Rubio-Tamayo, J.L.	The challenges of media and information literacy in the artificial intelligence ecology: deepfakes and misinformation | Retos de la Alfabetización Mediática e

		Informacional en la ecología de la Inteligencia Artificial: deepfakes y desinformación Resumen
12	Lendvai, G.F., Gosztonyi, G.	Deepfake and disinformation – What can the law do about fake news created with deepfake? | Deepfake y desinformación –¿Qué puede hacer el derecho frente a las noticias falsas creadas por deepfake?
13	Kumar, N., Kundu, A.	SecureVision: Advanced Cybersecurity Deepfake Detection with Big Data Analytics
14	Al-Dulaimi, O.A.H.H., Kurnaz, S.	A Hybrid CNN-LSTM Approach for Precision Deepfake Image Detection Based on Transfer Learning
15	Sánchez Esparza, M., Palella Stracuzzi, S., Fernández Fernández, Á.	Impact of Artificial Intelligence on RTVE: Verification of fake videos and Deep fakes, content generation, and new professional profiles | Impacto de la Inteligencia Artificial en RTVE: verificación de vídeos falsos y deepfakes, generación de contenidos y nuevos perfiles profesionales
16	Hasan, H.R., Salah, K., Yaqoob, I., Jayaraman, R., Omar, M.	NFTs for combating deepfakes and fake metaverse digital contents
17	Hameleers, M., van der Meer, T.G.L.A., Dobber, T.	They Would Never Say Anything Like This! Reasons To Doubt Political Deepfakes
18	Barrientos-Báez, A., Otero, T.P., Renó, D.P.	Fake images, Real Effects. Deepfakes as Manifestations of Gender-based Political Violence | Imágenes falsas, efectos reales. Deepfakes como manifestaciones de la violencia política de género
19	Murár, P., Kubovics, M., Jurišová, V.	THE IMPACT OF BRAND-VOICE INTEGRATION AND ARTIFICIAL INTELLIGENCE ON SOCIAL MEDIA MARKETING
20	Kharvi, P.L.	Understanding the Impact of AI-Generated Deepfakes on Public Opinion, Political Discourse, and Personal Security in Social Media
21	Hill, G., Waddington, M., Qiu, L.	From pen to algorithm: optimizing legislation for the future with artificial intelligence

22	Liu, Q., Xue, Z., Liu, H., Liu, J.	Enhancing Deepfake Detection With Diversified Self-Blending Images and Residuals
23	Hameleers, M., van der Meer, T.G.L.A., Dobber, T.	You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media
24	Ahmad, W., Shahzad, S.A., Hashmi, A., Ali, I., Ghaffar, F.	ResViT: A Framework for Deepfake Videos Detection
25	José, F., García-Ull, G.-U.	DeepFakes: The Next Challenge in Fake News Detection
26	Wahl-Jorgensen, K., Carlson, M.	Conjecturing Fearful Futures: Journalistic Discourses on Deepfakes
27	Vaccari, C., Chadwick, A.	Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News
28	Hasan, H.R., Salah, K.	Combating Deepfake Videos Using Blockchain and Smart Contracts

Grafica 1

Productividad de autores en investigaciones sobre deepfakes e inteligencia artificial



La gráfica proporciona una representación cuantitativa del nivel de productividad académica de distintos autores en el ámbito de estudio sobre *deepfakes* e inteligencia artificial. Se observa que el investigador Hameleers M. destaca como el autor con mayor número de contribuciones, con un total de tres publicaciones, lo que sugiere una participación continua y activa en este campo de investigación.

En un segundo nivel se encuentran los autores Dobber T., Hasan H.R., Salah K. y Van der Meer T.G.L.A., cada uno con dos documentos publicados. Esta recurrencia indica un interés sostenido por parte de estos investigadores en la temática, lo cual podría reflejarse en líneas de investigación consolidadas o colaboraciones recurrentes entre ellos.

Finalmente, autores como Ahmad W., Al-Dulaimi O.A.H.H., Ali I., Anaam E.A. y Azam S. presentan una única publicación en el corpus analizado. Su participación puede interpretarse como una incursión puntual o emergente en el estudio de los deepfakes, lo cual no descarta su potencial contribución futura al desarrollo de este campo.

En términos generales, la visualización permite identificar tanto a los investigadores con mayor trayectoria en el tema como a aquellos que han realizado aportes recientes. Este tipo de análisis resulta útil para reconocer actores clave en la producción científica, así como para explorar posibles redes de colaboración o referentes académicos en estudios sobre desinformación digital, tecnologías de detección y alfabetización mediática.

Tabla 3

Frecuencia de publicaciones por país en estudios sobre deepfakes e inteligencia artificial

Country	Freq
CHINA	16
SPAIN	13
NETHERLANDS	9
AUSTRALIA	6
SLOVAKIA	6
UNITED ARAB EMIRATES	6
MALAYSIA	4
AUSTRIA	3
CANADA	3
JAPAN	3

La tabla refleja la distribución geográfica de las publicaciones académicas relacionadas con *deepfakes* e inteligencia artificial, destacando la frecuencia de aparición de cada país en la autoría de los estudios analizados.

Se evidencia que China encabeza la producción científica en este campo, con un total de 16 publicaciones, lo cual denota un liderazgo investigativo probablemente vinculado al desarrollo tecnológico acelerado que caracteriza al país. En segundo lugar, se encuentra España, con 13 documentos, lo que indica una fuerte presencia de centros de investigación o universidades activas en el análisis crítico y técnico de los *deepfakes*.

Países Bajos ocupa la tercera posición con 9 publicaciones, seguido por Australia, Eslovaquia y los Emiratos Árabes Unidos, cada uno con 6 aportes, lo que sugiere un interés creciente en estos contextos tanto en los aspectos técnicos como éticos del fenómeno. A su vez, Malasia contribuye con 4 estudios, mientras que Austria, Canadá y Japón registran tres publicaciones cada uno, consolidando una participación relevante, aunque más limitada en comparación con los países líderes.

En conjunto, esta tabla permite inferir que la investigación sobre *deepfakes* y tecnologías asociadas está siendo desarrollada de forma significativa tanto en regiones asiáticas como europeas, con aportes clave desde instituciones de diversas latitudes. Esto refleja el carácter global

del fenómeno y la necesidad compartida de comprender y afrontar sus implicaciones en distintos contextos sociopolíticos y comunicativos.

Grafica 2

Nube de términos clave en investigaciones sobre deepfakes, redes sociales y desinformación

La gráfica muestra una nube de palabras que resalta los términos más frecuentes en investigaciones sobre *deepfakes*. Destacan palabras como "deepfake", "social media" y "deep learning", lo que refleja el interés académico en la manipulación digital y su difusión en plataformas sociales.

Términos como "misinformation", "disinformation" y "privacy" evidencian una preocupación por los riesgos sociales y éticos asociados, mientras que otros como "image forensics" y "cybersecurity" indican un enfoque técnico en la detección y prevención. En conjunto, la nube sintetiza las principales temáticas que conforman el debate actual sobre los impactos tecnológicos y sociales de los contenidos falsificados.

Tabla 4

Evolución anual de publicaciones sobre deepfakes (2019–2025)

Year	Articles
2019	1
2020	1
2021	2
2022	2
2023	0
2024	13
2025	9

La tabla presenta la evolución en el número de artículos publicados sobre *deepfakes* entre los años 2019 y 2025. Se observa un crecimiento lento y constante entre 2019 y 2022, con entre uno y dos artículos por año. Sin embargo, a partir de 2024 se evidencia un aumento significativo, con 13 publicaciones en ese año, seguido de 9 en 2025.

Este incremento notable sugiere que el interés académico y científico sobre esta temática ha crecido de forma acelerada en los últimos años, posiblemente impulsado por avances tecnológicos en inteligencia artificial, el aumento de casos de desinformación mediática y la necesidad de desarrollar marcos éticos y legales ante esta amenaza emergente. La ausencia de artículos en 2023 podría deberse a un desfase en la indexación o a un cambio en las bases de datos consultadas. En conjunto, la tendencia indica que el estudio de los *deepfakes* se ha consolidado como un campo de investigación emergente y de alta relevancia.

DISCUSIÓN

Los resultados de esta revisión sistemática evidencian que el desarrollo y proliferación de deepfakes, facilitados principalmente por el avance de las Redes Generativas Antagónicas (GANs) y el aprendizaje profundo, representan un desafío significativo para la confianza pública y la credibilidad de los medios de comunicación. Si bien estas tecnologías han mostrado potencial para aplicaciones legítimas en sectores como el entretenimiento y la educación, la literatura analizada coincide en que su uso malintencionado ha incrementado la vulnerabilidad de la sociedad frente a la desinformación y la manipulación mediática.

Diversos estudios señalan que la capacidad de los deepfakes para crear contenido audiovisual altamente realista dificulta la distinción entre información verídica y manipulada, lo que contribuye a una creciente desconfianza hacia los medios de comunicación tradicionales y digitales. Esta erosión de la confianza pública se traduce en una mayor susceptibilidad a la desinformación, ya que los ciudadanos tienden a cuestionar la autenticidad de los contenidos, incluso cuando estos provienen de fuentes reconocidas.

Uno de los ejemplos mas relevantes que se detectaron fueron de figuras políticas (Barack Obama, Donald Trump, Mark Zuckerberg). En el caso de Barack Obama en el 2018, el cineasta de Hollywood Jordan Peele creó un deepfake del expresidente estadounidense Barack Obama. En este vídeo, Obama hablaba sobre los peligros de las noticias falsas y se burlaba del entonces presidente Donald Trump. Este ejemplo buscaba denunciar los riesgos de esta tecnología. Por otro lado, para Donald Trump en el mismo año se mostró a Trump ofreciendo consejos a la gente de Bélgica sobre el cambio climático. Este vídeo fue creado por un partido político belga (el Partido Socialista Flamenco, SP.A) con el objetivo de generar un debate social. Al final del vídeo, Trump decía: "Todos sabemos que el cambio climático es falso, al igual que este vídeo", pero la última frase no se tradujo en los subtítulos holandeses, lo que generó indignación por la supuesta intromisión del presidente estadounidense en la política belga.

En lo que concierne a Mark Zuckerberg en junio de 2019, dos artistas británicos elaboraron un deepfake de alta calidad del CEO de Facebook, Mark Zuckerberg, que acumuló millones de visitas. El vídeo mostraba falsamente a Zuckerberg elogiando a "Spectre", una organización ficticia malvada de la serie de James Bond. Este vídeo, creado con imágenes de noticias, IA y un actor de voz, tenía como propósito demostrar cómo estas representaciones sintéticas pueden utilizarse para manipular la realidad.

Por otro lado, la credibilidad de los medios de comunicación se ve directamente amenazada por la circulación de deepfakes no detectados. La posibilidad de que los medios difundan inadvertidamente material manipulado puede afectar negativamente su reputación y la percepción de imparcialidad y rigor periodístico. En este sentido, la revisión destaca la necesidad de fortalecer los procesos de verificación y de invertir en tecnologías de detección, aunque los estudios coinciden en que los sistemas actuales no han logrado mantenerse al ritmo del desarrollo de nuevas técnicas de generación de deepfakes.

Adicionalmente, los riesgos asociados a los deepfakes, como la propagación de noticias falsas, el fraude, la suplantación de identidad y la manipulación política, plantean desafíos éticos y legales que requieren respuestas coordinadas entre el sector tecnológico, los medios de comunicación y los organismos reguladores. La literatura revisada enfatiza la importancia de la alfabetización mediática y digital como estrategia fundamental para mitigar el impacto negativo de los deepfakes en la opinión pública.

CONCLUSIONES

El análisis revela que la proliferación de deepfakes afecta la credibilidad de los medios al exponerlos al riesgo de difundir sin intención material manipulado, lo que disminuye la confiabilidad institucional y fortalece el escepticismo ciudadano respecto a la información consumida. Adicionalmente, el fenómeno genera efectos como la polarización social, el agotamiento del pensamiento crítico y la propagación de desinformación, especialmente en contextos electorales o ante figuras públicas.

Si bien existen aplicaciones benignas de esta tecnología, su uso malicioso exige respuestas integrales. La literatura destaca la urgencia de invertir en tecnologías de detección automatizada, fortalecer los marcos regulatorios con énfasis en la transparencia y educar a la ciudadanía en alfabetización mediática para reducir la vulnerabilidad social ante la manipulación digital.

La investigación académica sobre deepfakes ha mostrado un crecimiento significativo desde 2024, con China y España liderando la producción científica, reflejando el carácter global del fenómeno y la necesidad compartida de abordar sus implicaciones. En síntesis, la lucha contra los deepfakes es un imperativo social que demanda un esfuerzo coordinado y continuo entre la tecnología, la regulación y la educación para salvaguardar la integridad informativa y democrática.

REFERENCIAS

Ahmad, W., Ali, I., Adil Shahzad, S., Hashmi, A., & Ghaffar, F. (2022). ResViT: A Framework for Deepfake Videos Detection. *International Journal of Electrical and Computer Engineering Systems*, *13*(9), 807-813. https://doi.org/10.32985/ijeces.13.9.9

Al-Dulaimi, O. A. H. H., & Kurnaz, S. (2024). A Hybrid CNN-LSTM Approach for Precision Deepfake Image Detection Based on Transfer Learning. *Electronics*, *13*(9), 1662. https://doi.org/10.3390/electronics13091662

Barrientos-Báez, A., Piñeiro Otero, M. T., & Porto Renó, D. (2024). Imágenes falsas, efectos reales. Deepfakes como manifestaciones de la violencia política de género. *Revista Latina de Comunicación Social*, *82*, 1-30. https://doi.org/10.4185/rlcs-2024-2278

García-Ull, F. J. (2021a). «Deepfakes»: El próximo reto en la detección de noticias falsas. *Anàlisi*, 103-120. https://doi.org/10.5565/rev/analisi.3378

Hameleers, M., Van Der Meer, T. G. L. A., & Dobber, T. (2022). You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media. *Social Media + Society*, *8*(3). https://doi.org/10.1177/20563051221116346

Hameleers, M., Van Der Meer, T. G. L. A., & Dobber, T. (2024). They Would Never Say Anything Like This! Reasons To Doubt Political Deepfakes. *European Journal of Communication*, *39*(1), 56-70. https://doi.org/10.1177/02673231231184703

Hameleers, M., Van Der Meer, T., & Vliegenthart, R. (2025). How persuasive are political cheapfakes disseminated via social media? The effects of out-of-context visual disinformation on message credibility and issue agreement. *Information, Communication & Society*, *28*(1), 61-78. https://doi.org/10.1080/1369118x.2024.2388079

Hasan, H. R., & Salah, K. (2019). Combating Deepfake Videos Using Blockchain and Smart Contracts. *IEEE Access*, *7*, 41596-41606. https://doi.org/10.1109/access.2019.2905689

Hasan, H. R., Salah, K., Jayaraman, R., Yaqoob, I., & Omar, M. (2024). NFTs for combating deepfakes and fake metaverse digital contents. *Internet of Things*, *25*, 101133. https://doi.org/10.1016/j.iot.2024.101133

Hill, G., Waddington, M., & Qiu, L. (2025). From pen to algorithm: Optimizing legislation for the future with artificial intelligence. *AI & SOCIETY*, *40*(4), 3075-3086. https://doi.org/10.1007/s00146-024-02062-3

Khan, A. A., Laghari, A. A., Inam, S. A., Ullah, S., Shahzad, M., & Syed, D. (2025). A survey on multimedia-enabled deepfake detection: State-of-the-art tools and techniques, emerging trends, current challenges & limitations, and future directions. *Discover Computing*, *28*(1). https://doi.org/10.1007/s10791-025-09550-0

Kharvi, P. L. (2024). *Understanding the Impact of AI-Generated Deepfakes on Public Opinion, Political Discourse, and Personal Security in Social Media*.

Kumar, N., & Kundu, A. (2024). SecureVision: Advanced Cybersecurity Deepfake Detection with Big Data Analytics. *Sensors*, *24*(19), 6300. https://doi.org/10.3390/s24196300

Lee, F. Y., Kumaresan, S. P., Abdulwahab Anaam, E., & Chee Kong, W. (2025). Social Platforms in the Deepfake Age: Navigating Media Trust through Media Literacy. *JOIV : International Journal on Informatics Visualization*, *9*(1), 176. https://doi.org/10.62527/joiv.9.1.3490

Lendvai, G. F., & Gosztonyi, G. (2024a). Deepfake y desinformación –¿Qué puede hacer el derecho frente a las noticias falsas creadas por deepfake? *IDP. Revista de Internet, Derecho y Política*, *0*(41). https://doi.org/10.7238/idp.v0i41.427515

Lendvai, G. F., & Gosztonyi, G. (2024b). Deepfake y desinformación –¿Qué puede hacer el derecho frente a las noticias falsas creadas por deepfake? *IDP. Revista de Internet, Derecho y Política*, *0*(41). https://doi.org/10.7238/idp.v0i41.427515

Liu, Q., Xue, Z., Liu, H., & Liu, J. (2024a). Enhancing Deepfake Detection With Diversified Self-Blending Images and Residuals. *IEEE Access*, *12*, 46109-46117. https://doi.org/10.1109/access.2024.3382196

Liu, Q., Xue, Z., Liu, H., & Liu, J. (2024b). Enhancing Deepfake Detection With Diversified Self-Blending Images and Residuals. *IEEE Access*, *12*, 46109-46117. https://doi.org/10.1109/access.2024.3382196

Momeni, M. (2025). Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions Through Misinformation. *Journal of Creative Communications*, *20*(1), 41-56. https://doi.org/10.1177/09732586241277335

Murár, P., Kubovics, M., & Jurišová, V. (2024). The Impact of Brand-Voice Integration and Artificial Intelligence on Social Media Marketing. *Communication Today*, 50-63. https://doi.org/10.34135/communicationtoday.2024.vol.15.no.1.4

Ölvecký, M., Huraj, L., & Brlej, I. (2025). Evaluating the Effectiveness of Deepfake Video Detection Tools: A Comparative Study. *TEM Journal*, 64-77. https://doi.org/10.18421/tem141-07

Rubio Núñez, R. (2025). El uso de la inteligencia artificial en las campañas electorales y sus efectos democráticos. *Revista de Derecho Político*, *122*, 65-102. https://doi.org/10.5944/rdp.122.2025.44742

Sanchez-Acedo, A., Carbonell-Alcocer, A., Gertrudix, M., & Rubio-Tamayo, J.-L. (2024). The challenges of media and information literacy in the artificial intelligence ecology: Deepfakes and misinformation. *Communication & Society*, 223-239. https://doi.org/10.15581/003.37.4.223-239

Shibuya, Y., Nakazato, T., & Takagi, S. (2025a). How do people evaluate the accuracy of video posts when a warning indicates they were generated by AI? *International Journal of Human-Computer Studies*, *199*, 103485. https://doi.org/10.1016/j.ijhcs.2025.103485

Shibuya, Y., Nakazato, T., & Takagi, S. (2025b). How do people evaluate the accuracy of video posts when a warning indicates they were generated by AI? *International Journal of Human-Computer Studies*, *199*, 103485. https://doi.org/10.1016/j.ijhcs.2025.103485

Stracuzzi, S. P. (2024). *Impacto de la Inteligencia Artificial en RTVE: verificación de vídeos falsos y deepfakes, generación de contenidos y nuevos perfiles profesionales*.

Tran, V. H., Sebastian, Y., Karim, A., & Azam, S. (2024a). Distinguishing Human Journalists from Artificial Storytellers Through Stylistic Fingerprints. *Computers*, *13*(12), 328. https://doi.org/10.3390/computers13120328

Tran, V. H., Sebastian, Y., Karim, A., & Azam, S. (2024b). Distinguishing Human Journalists from Artificial Storytellers Through Stylistic Fingerprints. *Computers*, *13*(12), 328. https://doi.org/10.3390/computers13120328

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

Weikmann, T., Greber, H., & Nikolaou, A. (2025). After Deception: How Falling for a Deepfake Affects the Way We See, Hear, and Experience Media. *The International Journal of Press/Politics*, *30*(1), 187-210. https://doi.org/10.1177/19401612241233539

Xu, Z., Wen, X., Zhong, G., & Fang, Q. (2025). Public perception towards deepfake through topic modelling and sentiment analysis of social media data. *Social Network Analysis and Mining*, *15*(1). https://doi.org/10.1007/s13278-025-01445-8