Revista Latina de Comunicación Social
ISSN 1138-5820 / 85. 1-24.
https://doi.org/10.4185/rlcs-2027-2740

Percepciones del alumnado universitario sobre los *deepfakes* generados mediante inteligencia artificial: El caso de Grok

University students' perceptions of ai-generated *deepfakes*:
The case of Grok

Marta Sánchez-Hunt
Universidad de Cádiz
martahunt@uca.es

Estefanía Cestino González
Universidad de Málaga
ecestino@uma.es

Aránzazu Román-San-Miguel
Universidad de Sevilla
arantxa@us.es

Cómo citar este artículo / Referencia normalizada:

Sánchez-Hunt, Marta; Cestino González, Estefanía y Román-San-Miguel, Aránzazu (2027). Percepciones del alumnado universitario sobre los deepfakes generados mediante inteligencia artificial: El caso de Grok [University students' perceptions of ai-generated deepfakes: The case of Grok]. *Revista Latina de Comunicación Social*, 85, 1-24. https://doi.org/10.4185/RLCS-2027-2740

Fecha de Recepción: 31/03/2026
Fecha de Aceptación: 16/07/2026
Publicación Anticipada: 10/09/2026
Fecha de Publicación: 01/01/2027

RESUMEN

Introducción: En el contexto de expansión de la Inteligencia Artificial (IA) generativa, los *deepfakes* emergen como un problema específico que afecta a la veracidad informativa, la identidad y la confianza social. La accesibilidad de herramientas como Grok intensifica estos desafíos, especialmente entre el alumnado universitario, un colectivo clave en la alfabetización digital. El estudio tiene como objetivo analizar sus

percepciones sobre los riesgos, la regulación y las necesidades formativas asociadas a esta tecnología. **Metodología:** Se adopta un enfoque cuantitativo exploratorio basado en un cuestionario estructurado. La muestra está compuesta por 249 estudiantes de tres instituciones andaluzas: Universidad de Sevilla (US), Universidad de Málaga (UMA) y Centro Universitario EUSA. El instrumento recoge percepciones sobre riesgos, regulación y formación en relación con los *deepfakes* generados mediante Grok. **Resultados:** La desinformación aparece como el principal riesgo percibido, seguida de la suplantación de identidad y la vulneración de la privacidad. Además, existe una demanda mayoritaria de regulaciones estrictas, mayor transparencia algorítmica y programas avanzados de alfabetización mediática. Grok es valorado de forma predominantemente crítica. **Discusión:** Los resultados reflejan una preocupación significativa por el impacto social de los *deepfakes* y evidencian una percepción crítica hacia su uso. La demanda de regulación y formación indica una conciencia creciente sobre los riesgos, así como la necesidad de herramientas para afrontarlos. **Conclusiones:** Los *deepfakes* presentan una naturaleza ambivalente que requiere marcos normativos proactivos, responsabilidad ética compartida y una formación universitaria orientada a la gobernanza crítica de la IA.

Palabras clave: *deepfakes*; desinformación; ética digital; Grok; inteligencia artificial; universitarios.

ABSTRACT

Introduction: In the current context of expanding generative Artificial Intelligence, *deepfakes* emerge as a specific challenge affecting information accuracy, identity, and social trust. The accessibility of tools such as Grok intensifies these issues, particularly among university students, a key group in digital literacy processes. This study aims to analyze their perceptions regarding risks, regulation, and training needs associated with this technology. **Methodology:** A quantitative exploratory approach was adopted using a structured questionnaire. The sample consisted of 249 students from three Andalusian institutions: the University of Seville, the University of Málaga, and Centro Universitario EUSA. The instrument collected perceptions on risks, regulation, and educational needs related to *deepfakes* generated through Grok. **Results:** Misinformation is identified as the main perceived risk, followed by identity impersonation and privacy violations. There is also a strong demand for stricter regulatory frameworks, greater algorithmic transparency, and advanced media literacy programs. Grok is predominantly viewed in a critical light. **Discussion:** The findings reveal significant concern about the social impact of *deepfakes* and a critical perception of their use. The demand for regulation and training reflects growing awareness of the risks and the need for tools to address them effectively. **Conclusions:** *Deepfakes* have an ambivalent nature that requires proactive regulatory frameworks, shared ethical responsibility, and university education focused on the critical governance of artificial intelligence.

Keywords: artificial intelligence; *deepfakes*; digital ethics; disinformation; Grok; university students.

1. INTRODUCCIÓN

El desarrollo acelerado de la inteligencia artificial (IA) generativa ha transformado de forma sustancial la producción y circulación de contenidos audiovisuales, situando a los *deepfakes* en el centro del debate académico y social por su capacidad para alterar la percepción de la realidad y desafiar las nociones de veracidad y autenticidad (Chesney y Citron, 2019). Concretamente en contextos educativos, las tecnologías *deepfake* que utilizan inteligencia artificial generativa han experimentado un interés importante. Roe *et al*. (2025) concluyeron en un estudio sobre esta realidad que la intención de usar *deepfakes* era relativamente baja entre el profesorado ya que, aunque les podían ayudar a atraer más la atención de su alumnado, también implicaba riesgos; por lo que sostienen que es necesaria una reflexión profunda sobre su uso para mantener la integridad y la autonomía profesionales. La incorporación de herramientas accesibles como Grok, que permiten la generación de vídeos hiperrealistas manipulados mediante técnicas de *deep learning* sin filtros éticos explícitos, intensifica las preocupaciones relacionadas con la desinformación, la suplantación de identidad y los

usos fraudulentos, especialmente en ecosistemas digitales polarizados (Vaccari y Chadwick, 2020).

Al mismo tiempo, la literatura reciente reconoce el carácter ambivalente de los *deepfakes*, al señalar tanto sus riesgos comunicativos como su potencial creativo e innovador en ámbitos como la comunicación, la publicidad o la educación, siempre que su uso se integre en entornos controlados y con supervisión ética (Dayal *et al*., 2024; Koenecke *et al*., 2020; Sundar *et al*., 2021). Esta dualidad plantea interrogantes éticos y regulatorios que requieren enfoques multidisciplinares y estrategias orientadas a reforzar la alfabetización mediática y la transparencia tecnológica (Carlson, 2015; Peña-Fernández *et al*., 2023; Porlezza y Schapals, 2024). Algunos estudios ya hablan de la transparencia de la IA, como del de Bustos Díaz y Martín-Vicario (2024), quienes aseguran que ChatGPT indica las fuentes utilizadas en la búsqueda y estas suelen ser sólidas, validando así la transparencia de herramientas de IA. Otros, como Guevara Chávez y Almeida Macias (2025), han identificado las estrategias más eficaces para neutralizar los riesgos de la información automatizada y concluyen que la formación crítica, la verificación automatizada y la educación algorítmica son las estrategias más efectivas. Consideran que se debe adaptar la alfabetización mediática tradicional para hacer frente a las amenazas que puedan suscitar los bots, los filtros algorítmicos o los *deepfakes*. Como se aprecia en las investigaciones actuales, en este momento se está estudiando la desinformación en relación con los retos que está planteando la IA y así lo constatan Torres-Toukoumidis *et al*. (2025) en una revisión bibliométrica realizada entre 2013 y 2025 sobre IA y alfabetización mediática.

Pese al creciente interés académico, la mayor parte de los estudios se han centrado en aproximaciones técnicas o normativas, existiendo un menor número de investigaciones empíricas que analicen cómo estudiantes universitarios y expertos perciben el uso de *deepfakes* cuando se integran en plataformas accesibles y ampliamente difundidas como Grok. Atendiendo a esta laguna, el presente estudio analiza, desde una perspectiva exploratoria, las percepciones del alumnado universitario sobre los *deepfakes* generados mediante inteligencia artificial. De forma complementaria, incorpora la valoración de expertos en comunicación y ética mediática con el fin de contextualizar e interpretar los resultados obtenidos.

2. ESTADO DE LA CUESTIÓN

2.1. Orígenes y definición de los *deepfakes*

El término *deepfake* comenzó a utilizarse a partir de 2017 para describir contenidos audiovisuales manipulados mediante técnicas de *deep learning*, marcando una nueva etapa en la alteración digital de imágenes y vídeos caracterizada por un alto grado de realismo y una amplia capacidad de difusión (Ballesteros-Aguayo y Ruiz del Olmo, 2024; Bañuelos Capistrán, 2022). Aunque la manipulación visual cuenta con precedentes históricos, el desarrollo reciente de los *deepfakes* supone un salto cualitativo debido a su accesibilidad tecnológica, su sofisticación semántica y su rápida expansión en entornos digitales.

La literatura académica coincide en señalar la ausencia de una definición universalmente consensuada de *deepfake*, lo que ha generado una notable heterogeneidad conceptual. Mientras algunos estudios restringen el término a vídeos manipulados, otros amplían su alcance para incluir imágenes estáticas, audios sintéticos y otros contenidos generados mediante inteligencia artificial, dificultando la comparabilidad entre investigaciones y la formulación de marcos regulatorios coherentes (Vasist y Krishnan, 2022; Whittaker *et al*., 2023). Esta ambigüedad terminológica se refleja también en el ámbito normativo, donde persisten discrepancias en la categorización jurídica de estas tecnologías.

Pese a esta diversidad conceptual, existe un consenso técnico básico en definir los *deepfakes* como contenidos audiovisuales sintéticos generados mediante redes neuronales profundas capaces de producir sustituciones

faciales, clonación de voz o recreaciones corporales con un elevado nivel de verosimilitud (Ballesteros-Aguayo y Ruiz del Olmo, 2024; Maniyal y Kumar, 2024). Diversos estudios subrayan que estas características convierten a los *deepfakes* en un fenómeno de alcance global, con implicaciones relevantes para la identidad, la privacidad y la confianza en la información audiovisual (Gomes-Gonçalves, 2022).

2.2. Impacto positivo: creatividad en innovación

Aunque una parte significativa de la literatura se ha centrado en los riesgos asociados a los *deepfakes*, investigaciones recientes subrayan su potencial creativo, educativo y cultural, destacando su capacidad para generar innovación en distintos ámbitos cuando se integran en entornos controlados y con supervisión ética. En el contexto educativo y cultural, el uso de contenidos sintéticos puede favorecer el desarrollo del pensamiento crítico, la reflexión ética y nuevas formas de experimentación visual, al permitir analizar simultáneamente sus aplicaciones creativas y sus implicaciones sociales (Murillo Ligorred y Ramos Vallecillo, 2025; Murillo-Ligorred *et al*., 2023; Okhrymovych, 2024).

Desde una perspectiva más amplia, los *deepfakes* contribuyen a la reconfiguración de los conceptos tradicionales de creatividad, autenticidad e identidad en la cultura digital, ampliando las posibilidades narrativas y expresivas en sectores como la producción audiovisual, las industrias culturales y la comunicación visual (Cejudo, 2024; Kalpokas y Kalpokiene, 2022). Estas tecnologías permiten explorar nuevas formas de creación simbólica y participación cultural, democratizando procesos creativos que anteriormente requerían altos niveles de especialización técnica. Asimismo, investigaciones recientes apuntan a la aplicación de los *deepfakes* en ámbitos no artísticos, como la formación profesional, la simulación o el aprendizaje avanzado, donde pueden contribuir a mejorar experiencias educativas y metodologías innovadoras (A. Kaur *et al*., 2024; Nannaware *et al*., 2025). Así como en sectores como el entretenimiento, la publicidad, los medios de comunicación y la producción audiovisual, donde su uso se ha expandido significativamente en los últimos años (Kılıç y Kahraman, 2023). En conjunto, estos trabajos coinciden en señalar que los *deepfakes* han dejado de ser un fenómeno exclusivamente asociado a la desinformación para convertirse también en herramientas con capacidad de impulsar innovación creativa y educativa en la sociedad digital, siempre que su uso esté acompañado de criterios éticos y marcos de gobernanza adecuados.

2.3. Riesgos: ética, desinformación y privacidad

Los *deepfakes* plantean riesgos significativos en los ámbitos ético, informativo y de privacidad, al intensificar la crisis contemporánea de veracidad y erosionar la confianza pública en los contenidos audiovisuales. Estas tecnologías comprometen principios fundamentales como la veracidad, el consentimiento y la autodeterminación, al facilitar la creación y difusión de contenidos manipulados altamente verosímiles, difíciles de detectar y de atribuir a un responsable concreto (J. Kaur *et al*., 2024; Singh, 2024). En este sentido, diversos estudios subrayan la necesidad de desarrollar técnicas avanzadas de detección y marcos éticos robustos para mitigar los riesgos asociados a los *deepfakes* y la inteligencia artificial generativa (Gupta y Fatunmbi, 2024).

Desde esta perspectiva, los *deepfakes* generan dilemas éticos y legales complejos relacionados con la vulneración de derechos de la personalidad, la protección de la privacidad y la dificultad de establecer responsabilidades jurídicas en entornos digitales caracterizados por la automatización y la opacidad algorítmica (Chapagain *et al*., 2024). Estas limitaciones normativas, aun insuficientemente resueltas, amplifican los riesgos de usos maliciosos y contribuyen a la expansión de prácticas como la suplantación de identidad, la manipulación emocional y la desinformación sistemática, con impactos negativos en la salud mental y la cohesión social.

Como señalan Ali *et al*. (2022) y Hasani *et al*. (2024) el impacto de los *deepfakes* resulta especialmente relevante en el ámbito político y democrático, donde su utilización puede favorecer la manipulación de la opinión pública,

la alteración de procesos electorales y la difusión de campañas de desinformación a gran escala, con implicaciones para la seguridad institucional y la estabilidad social. Asimismo, la literatura subraya que estas tecnologías afectan al periodismo, a los derechos humanos y a la protección de colectivos vulnerables, al dificultar la verificación de contenidos y facilitar la producción de imágenes o audios no consentidos, especialmente de carácter sexual (Gregory, 2022). En conjunto, los estudios revisados muestran que los riesgos asociados a los *deepfakes* son multidimensionales y requieren respuestas integrales basadas en marcos regulatorios específicos, alfabetización mediática avanzada y el desarrollo de infraestructuras de autenticidad que refuercen la protección de la ciudadanía en la era de la inteligencia artificial generativa.

2.4. Grok como facilitador de *deepfakes*

Grok, como modelo generativo de nueva generación integrado en plataformas de alta circulación como X, desempeña un papel relevante en la expansión y sofisticación de los *deepfakes*. Desde una perspectiva comunicativa, y siguiendo a Bañuelos Capistrán (2020), el *deepfake* puede entenderse como una especie mediática emergente inscrita en la lógica de la posverdad, en la que los contenidos audiovisuales dejan de ser meros reflejos de la realidad para convertirse en agentes activos de su construcción. La combinación entre generación automatizada, viralidad e inmediatez sitúa así a Grok como un elemento clave dentro del ecosistema contemporáneo de la desinformación digital.

Los sistemas como Grok no solo generan contenidos sintéticos, sino que pueden intervenir directamente en el flujo informativo en tiempo real, llegando incluso a operar como pseudo-verificadores de hechos, lo que introduce riesgos epistémicos significativos en contextos educativos y sociales (Yamaguchi *et al*., 2025). En el ámbito universitario, investigaciones cualitativas han detectado que el uso intensivo de herramientas de IA generativa se asocia a déficits en pensamiento crítico, competencias de evaluación informacional y comprensión de los límites éticos de estos sistemas, lo que amplifica las vulnerabilidades cognitivas de sus usuarios (Caballero Mariscal *et al*., 2025).

El impacto de Grok se manifiesta también en su capacidad para facilitar la producción de narrativas falsas con apariencia de legitimidad académica o científica. Acquier y Cossey (2025) introducen el concepto de *academic deepfake* para describir la generación de documentos pseudocientíficos atribuidos falsamente a versiones experimentales del modelo, capaces de cuestionar consensos consolidados y de introducir duda estratégica en debates sensibles.

Otro ámbito problemático es el de la moderación algorítmica. Análisis críticos señalan que Grok presenta barreras de seguridad insuficientes, lo que incrementa la probabilidad de generación o difusión de contenidos manipuladores o dañinos, incluso cuando se incorporan mecanismos correctivos basados en moderación comunitaria (Abrar, 2025). Estas limitaciones estructurales evidencian la fragilidad del modelo en la gestión de contenidos sensibles dentro de plataformas abiertas.

La accesibilidad técnica de Grok ha sido vinculada asimismo a riesgos graves en materia de explotación digital. Estudios recientes muestran que modelos avanzados reducen de forma drástica las barreras para la creación de identidades sintéticas, la clonación de voces y la producción de material falso utilizado con fines de control, extorsión o encubrimiento de actividades ilícitas, incluyendo la explotación sexual digital (Levesque, 2025; Kamachee *et al*., 2025).

En el ámbito corporativo y jurídico, la literatura advierte que la combinación entre *deepfakes* y herramientas generativas como Grok facilita estrategias avanzadas de *social engineering* y plantea desafíos significativos para la autenticación de evidencias audiovisuales, comprometiendo tanto la ciberseguridad organizacional como la

percepción de la verdad jurídica en procesos judiciales (Pedersen *et al*., 2025; Iannuzzi *et al*., 2025). De forma complementaria, se ha documentado el uso sistemático de modelos multimodales para la producción de desinformación a gran escala, aprovechando la capacidad de generar contenidos manipulados en múltiples formatos (Carpenter, 2024; Sapkota *et al*., 2025).

En conjunto, Grok no puede entenderse como una herramienta generativa neutral, sino como un acelerador estructural del ecosistema *deepfake*, cuyo impacto deriva de la convergencia entre accesibilidad, multimodalidad, integración social y capacidad de producción masiva de contenidos convincentes. Su expansión plantea retos sustantivos para la educación, la veracidad informativa, la seguridad digital, la regulación y la justicia, reforzando la necesidad de marcos éticos y de gobernanza capaces de mitigar sus efectos.

3. OBJETIVOS

Los objetivos que plantea este estudio sobre los *deepfakes* generados por Grok son tres:

1. Determinar el nivel de conocimiento y la percepción general del alumnado universitario sobre los *deepfakes* generados mediante Grok.

2. Complementar la percepción del alumnado universitario mediante la valoración de expertos en comunicación, ética mediática y tecnologías digitales, con el fin de contextualizar los riesgos, beneficios y desafíos regulatorios asociados a los *deepfakes* generados mediante inteligencia artificial.

3. Discernir a través del análisis de la muestra y la literatura al respecto, la necesidad de marcos regulatorios, así como de protocolos y prácticas responsables para la creación y uso ético de *deepfakes* en contextos educativos y sociales, atendiendo a aspectos éticos, legales y comunicativos.

4. METODOLOGÍA

La investigación adopta un enfoque cuantitativo exploratorio basado en una encuesta estructurada aplicada a estudiantes universitarios. De forma complementaria, se ha incorporado la valoración de expertos en comunicación, ética mediática y tecnologías digitales con el objetivo de contextualizar e interpretar los resultados obtenidos. Esta aportación tiene un carácter auxiliar y no constituye una fase cualitativa independiente del estudio.

Para la recogida de datos se ha empleado un cuestionario administrado telemáticamente mediante Google Forms. Antes de su aplicación, ha sido sometido a una revisión de contenido orientada a garantizar la claridad, pertinencia y adecuación de los ítems a los objetivos de la investigación. Este proceso permite revisar la formulación de las preguntas, evitar ambigüedades y asegurar la correspondencia entre los bloques temáticos del cuestionario y las dimensiones analizadas en el estudio.

En la página inicial del cuestionario enviado a la muestra se informaba a los participantes sobre los objetivos de la investigación, el carácter voluntario de la participación, la finalidad académica del estudio, el tratamiento confidencial de la información y la posibilidad de abandonar el cuestionario en cualquier momento. La continuación del formulario se consideró manifestación de consentimiento para participar en la investigación.

El instrumento se aplicó a una muestra de 249 estudiantes procedentes de la Universidad de Sevilla (US), la Universidad de Málaga (UMA) y el Centro Universitario EUSA. Los participantes pertenecían a titulaciones vinculadas al Periodismo, la Publicidad y Relaciones Públicas, el Marketing y la Economía y Administración y Dirección de Empresas, distribuidos entre segundo, tercer y cuarto curso. Esta diversidad académica permitió incorporar perspectivas procedentes de distintos ámbitos de la comunicación, la gestión empresarial y el

análisis de mercados. La selección de participantes se ha realizado mediante un muestreo no probabilístico por conveniencia, atendiendo a criterios de accesibilidad. En consecuencia, no es posible establecer márgenes de error ni niveles de confianza en términos estadísticos, por lo que los resultados deben interpretarse desde una perspectiva exploratoria.

El proceso de recogida de datos se llevó a cabo entre el 25 de septiembre y el 12 de diciembre de 2025, mediante la administración de un cuestionario online distribuido a través de canales institucionales y académicos. Durante el proceso de depuración de datos se eliminaron respuestas incompletas y duplicadas, obteniéndose un total válido de 249 cuestionarios analizados.

El instrumento de recogida de datos consistió en un cuestionario estructurado compuesto por 18 ítems distribuidos en cuatro bloques temáticos: consideraciones éticas (6 ítems), inteligencia artificial y responsabilidad (5 ítems), regulación y medidas preventivas (4 ítems) y perspectivas futuras (3 ítems). El cuestionario combinaba preguntas cerradas de respuesta única y categorías previamente definidas orientadas a identificar las percepciones, valoraciones y posicionamientos de los participantes respecto al fenómeno de los *deepfakes* y su relación con herramientas de inteligencia artificial generativa como Grok. Aunque Grok constituye el caso de referencia que motiva la investigación, varias de las preguntas del cuestionario fueron formuladas de manera general con el propósito de explorar percepciones aplicables al fenómeno de los *deepfakes* generados mediante inteligencia artificial en un sentido más amplio. Esta decisión ha permitido analizar no solo las valoraciones asociadas a una herramienta concreta, sino también las implicaciones éticas, sociales y regulatorias vinculadas a este tipo de tecnologías. Los resultados se analizaron mediante estadística descriptiva, calculando frecuencias y porcentajes por ítem en función del número total de respuestas válidas en cada caso.

La participación fue voluntaria. Aunque el cuestionario incluía determinados datos identificativos de carácter opcional, estos no fueron utilizados en el análisis de resultados ni en la elaboración de las conclusiones del estudio. La información recopilada fue tratada de forma confidencial y utilizada exclusivamente con fines académicos y de investigación. Así, el estudio se ha realizado conforme al Reglamento (UE) 2016/679 (RGPD) y a la Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales. La participación fue voluntaria y anónima, sin recogida de datos identificativos obligatorios. Los participantes fueron informados de la finalidad académica del estudio, del tratamiento confidencial de la información y del uso exclusivo de los datos con fines de investigación. Asimismo, se garantizó la no cesión de datos a terceros y su conservación únicamente durante el tiempo necesario para el análisis. Los participantes pudieron ejercer sus derechos de acceso, rectificación, supresión y limitación del tratamiento conforme a la normativa vigente.

Paralelamente, se administró también el cuestionario a 3 expertos seleccionados mediante muestreo intencional en función de su trayectoria en ética mediática, comunicación digital y creatividad audiovisual. Se incluyó al Dr. Juan Carlos Suárez Villegas, catedrático de Ética de la Comunicación en la Universidad de Sevilla, por su reconocida labor investigadora en deontología profesional, a la Dra. Sonia Blanco, profesora Titular de Comunicación en la Universidad de Málaga, por su experiencia en análisis digital y dinámicas comunicativas contemporáneas y a María del Rosario Marín Pinilla, experta en Comunicación Estratégica, consultora de Comunicación, y profesora de Publicidad y Relaciones Públicas del Centro Universitario EUSA Sevilla, por su especialización en creatividad publicitaria y producción audiovisual. La selección de un experto por cada universidad representada en el estudio del alumnado asegura equilibrio institucional y diversidad epistemológica, lo que permite contrastar la percepción estudiantil con valoraciones académicamente fundamentadas. Las aportaciones realizadas por los expertos fueron empleadas como elemento complementario para contextualizar e interpretar los resultados obtenidos en la encuesta aplicada al alumnado universitario, sin constituir una fase cualitativa independiente del diseño de investigación.

Ambos instrumentos fueron sometidos a revisión de contenido con el fin de garantizar la claridad, pertinencia y adecuación de los ítems al objeto de estudio. Los datos obtenidos del alumnado universitario fueron analizados mediante estadística descriptiva, utilizando frecuencias y porcentajes para identificar tendencias generales de percepción. Por su parte, las aportaciones realizadas por los expertos fueron empleadas como elemento complementario de contextualización e interpretación de los resultados obtenidos, permitiendo enriquecer la discusión desde una perspectiva especializada. Estas contribuciones no constituyen una fase cualitativa independiente ni un proceso formal de triangulación metodológica, sino un recurso de apoyo para la comprensión del fenómeno analizado.

Durante el proceso de elaboración del manuscrito se emplearon herramientas digitales y sistemas basados en inteligencia artificial como apoyo en distintas fases del estudio. En concreto, se utilizó ChatGPT (OpenAI) en fechas comprendidas entre el 15 diciembre de 2025 y el 10 de febrero de 2026, con fines de apoyo en la estructuración del texto y mejora de la claridad expositiva. No obstante, en ningún caso el texto de este artículo ha sido generado con inteligencia artificial. Asimismo, se recurrió a Google Scholar y Google Scholar Labs para la localización, exploración y selección de literatura científica relevante, garantizando el acceso a fuentes académicas actualizadas. Por otra parte, se empleó Perplexity AI como herramienta de apoyo en la exploración y síntesis inicial de información y en tareas auxiliares de organización de datos, sin que ello implicara la automatización del análisis empírico.

La recogida de datos se realizó mediante Google Forms, herramienta utilizada para la administración del cuestionario y la generación preliminar de gráficos descriptivos. En todos los casos, estas herramientas se utilizaron exclusivamente como apoyo técnico. El diseño metodológico, la validación de los datos, el análisis de resultados y la interpretación científica fueron realizados íntegramente por las autoras, quienes asumen la plena responsabilidad del contenido del estudio. En ningún caso estas herramientas sustituyeron el análisis científico ni la interpretación de los resultados, siendo la autoría intelectual, el diseño metodológico y la discusión crítica responsabilidad exclusiva de las autoras.

5. RESULTADOS

El análisis conjunto de las 249 respuestas del alumnado universitario revela una percepción ampliamente consolidada de los *deepfakes* como una de las tecnologías más disruptivas del actual ecosistema comunicativo, con un impacto directo sobre la veracidad informativa, la privacidad, la confianza social y la seguridad digital. De manera transversal, los participantes identifican a los *deepfakes* como un fenómeno en rápida expansión, técnicamente sofisticado y con una capacidad creciente para mimetizar la realidad. Las valoraciones aportadas por los expertos se utilizaron como elemento complementario para contextualizar e interpretar los resultados obtenidos.

Figura 1. *Riesgos éticos del uso de Deepfakes en la sociedad actual*

Fuente: Elaboración propia.

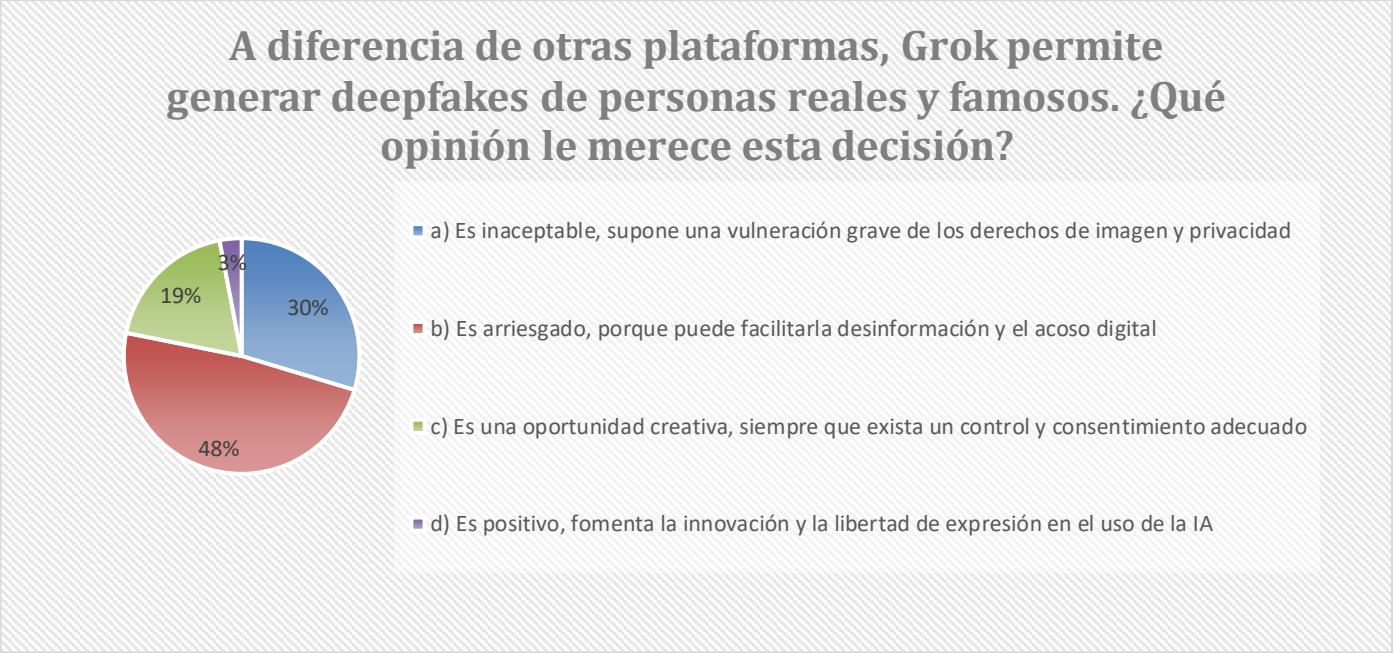
Los resultados representados en la figura 1 muestran que el principal riesgo asociado a los *deepfakes* es la desinformación, señalada como la amenaza dominante por cerca de la mitad de la muestra total. Este resultado confirma que el impacto más alarmante no se percibe únicamente en términos individuales, sino como un problema estructural de la esfera pública, vinculado a la manipulación de la opinión pública, la degradación de la credibilidad mediática y la dificultad progresiva para distinguir entre contenido real y sintético. Junto a la desinformación aparecen una serie de riesgos secundarios de alta relevancia como son la vulneración de la privacidad, la suplantación de identidad y el fraude digital, por lo que se configura un escenario de amenaza múltiple que afecta simultáneamente a ciudadanos, instituciones y empresas.

Figura 2. *Impacto de los Deepfakes en la confianza del público hacia los medios de comunicación y las redes sociales*

Fuente: Elaboración propia.

En relación con el impacto de los *deepfakes* sobre la confianza pública (figura 2), la mayoría de los encuestados considera que estos contenidos reducen de manera significativa la credibilidad de los medios de comunicación y de las redes sociales. Predomina la percepción de que la proliferación de vídeos, audios e imágenes sintéticas está erosionando progresivamente la capacidad del público para confiar en la autenticidad de los mensajes, lo que genera un contexto de escepticismo informativo estructural.

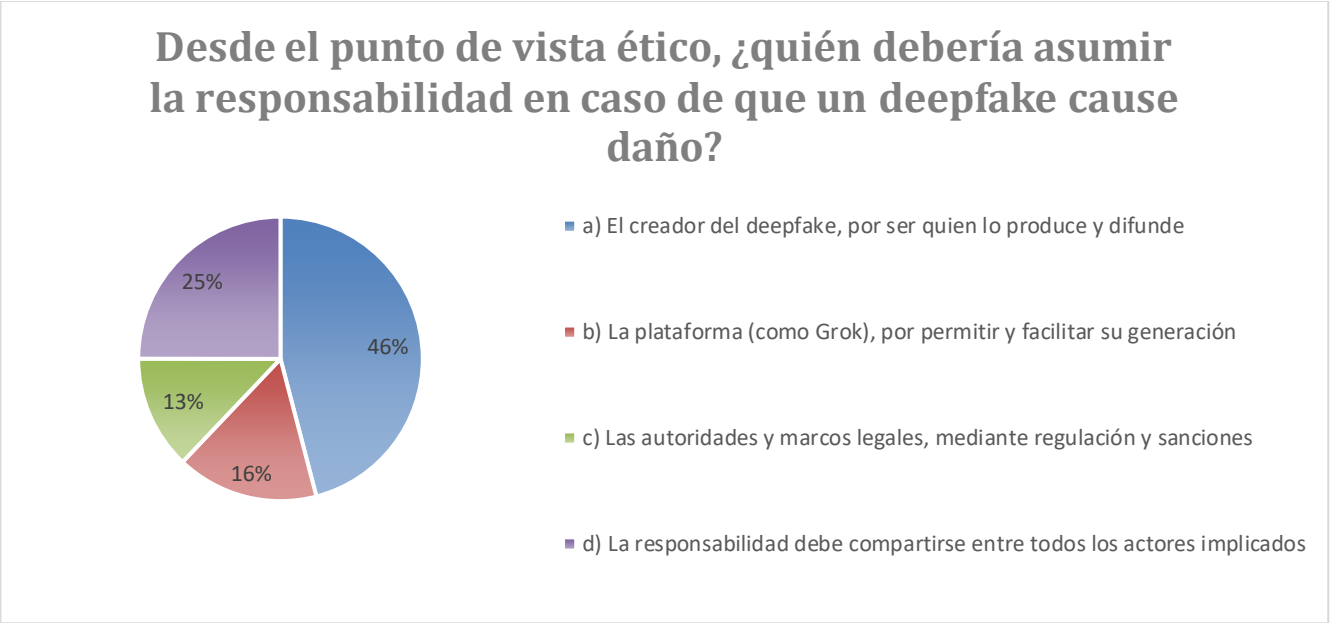
Figura 3. Opinión sobre la posibilidad de generar *deepfakes* con Grok



Fuente: Elaboración propia**.**

La valoración global de Grok como herramienta capaz de generar *deepfakes* es mayoritariamente crítica (figura 3). La mayoría de los participantes manifiesta una postura de preocupación ante la facilidad de acceso a tecnologías de generación sintética, especialmente cuando estas permiten reproducir la imagen y la voz de personas reales. Grok es percibido como un acelerador del problema, al reducir barreras técnicas y facilitar la producción de contenidos manipulados con apariencia de autenticidad. Esta percepción negativa no se basa únicamente en el potencial tecnológico, sino en su integración en plataformas de consumo masivo y alta viralidad.

Figura 4. *Quién debería asumir la responsabilidad ante los daños derivados de un deepfake*

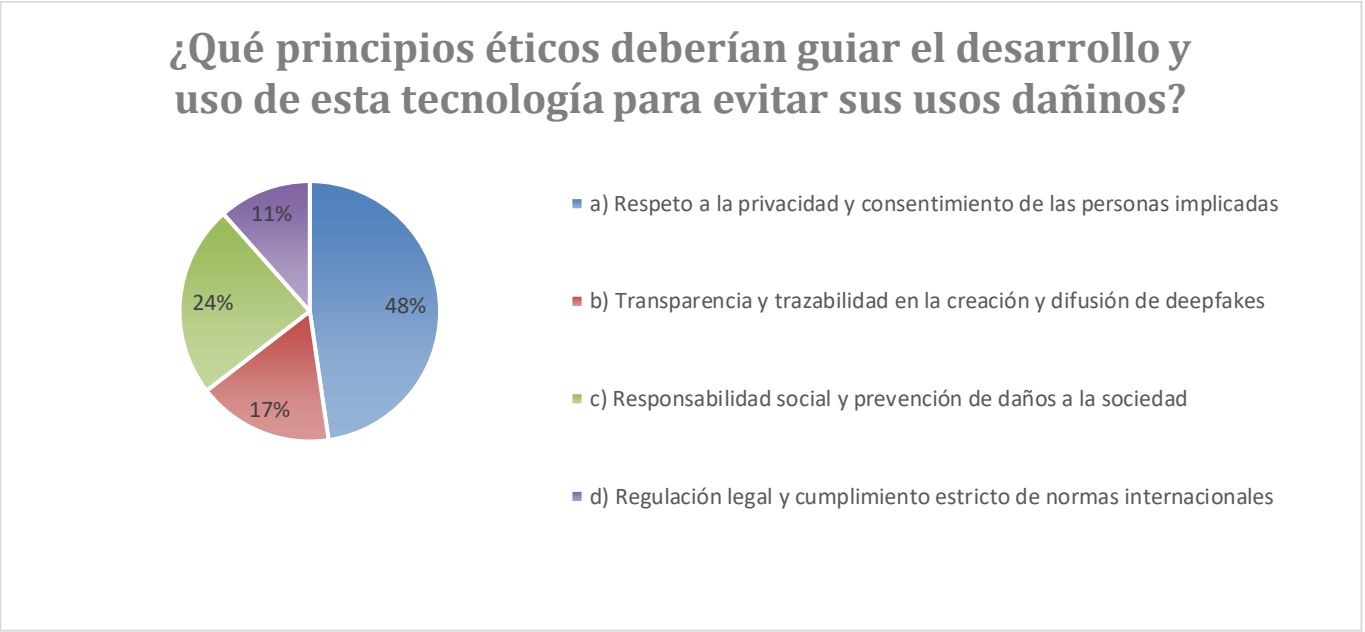


Fuente: Elaboración propia.

En cuanto a la responsabilidad ante los daños derivados de un *deepfake* (figura 4), los resultados muestran un consenso claro en atribuir la responsabilidad principal a quien crea y difunde el contenido manipulado. No obstante, de forma complementaria, una parte significativa de la muestra considera que las plataformas tecnológicas y los desarrolladores de sistemas de IA deben asumir una cuota relevante de corresponsabilidad, especialmente por su papel en la amplificación del contenido y en el diseño de herramientas con escasos filtros preventivos.

La demanda de regulación es uno de los hallazgos más sólidos y transversales del estudio. De manera mayoritaria, los participantes consideran que el simple etiquetado de los *deepfakes* como contenido sintético es insuficiente, y que resulta imprescindible establecer marcos normativos más estrictos, capaces de proteger la privacidad, la identidad personal, la reputación y la seguridad jurídica. En este sentido, se observa una clara orientación hacia modelos de regulación proactiva, que integren sanciones, mecanismos de detección y obligaciones de transparencia algorítmica.

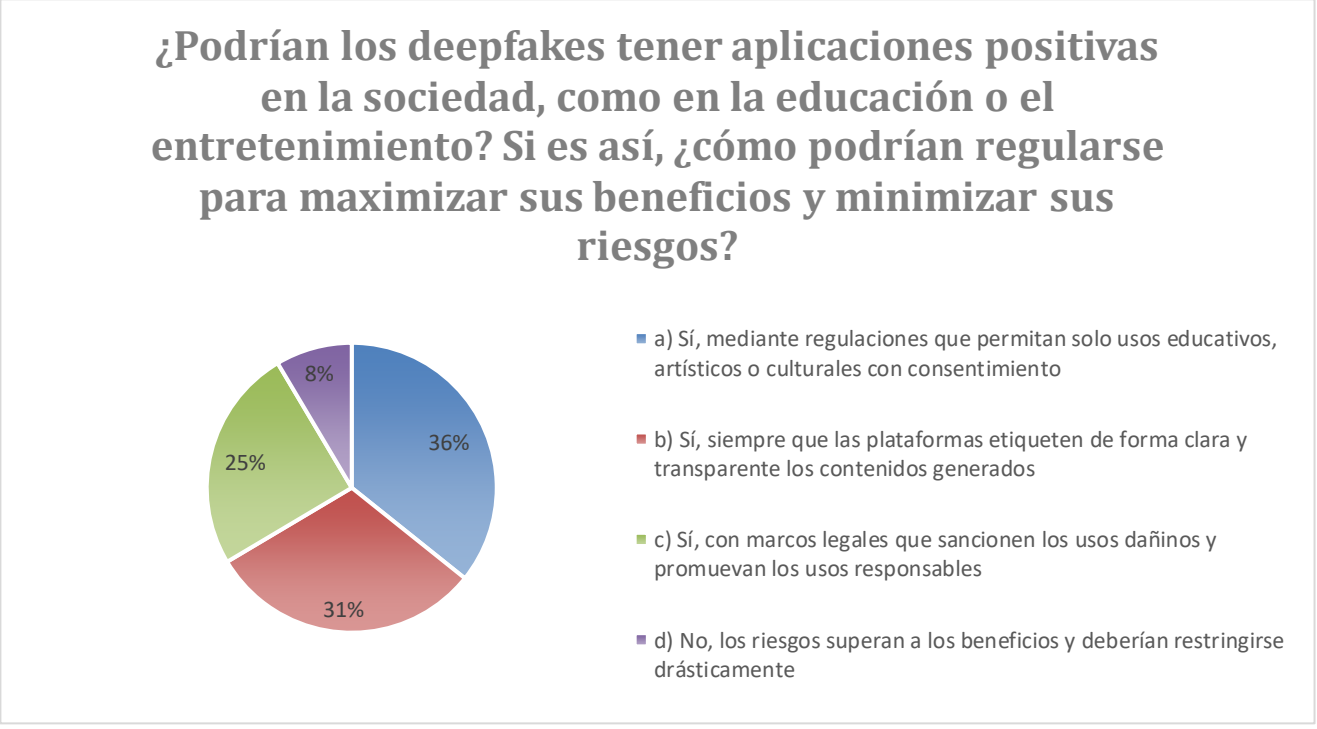
Figura 5. Principios éticos que deben guiar el desarrollo y uso de los *deepfakes*



Fuente: Elaboración propia.

En relación con los principios éticos que deben guiar el desarrollo y uso de los *deepfakes*, emergen de manera consistente cuatro ejes fundamentales: el respeto a la privacidad, el consentimiento explícito, la transparencia en la generación de contenidos sintéticos y la prevención del daño social (figura 5). Estos principios funcionan como un marco moral compartido entre el conjunto de la muestra, independientemente del perfil académico o institucional. (Esta pregunta obtuvo 244 respuestas válidas).

Figura 6. Posibilidad de un uso positivo de los *deepfakes y cómo podrían regularse*



Fuente: Elaboración propia.

Respecto a los usos positivos de los *deepfakes* (figura 6), los resultados globales muestran una postura condicionada: una parte relevante del alumnado reconoce que estas tecnologías pueden tener aplicaciones legítimas en ámbitos como la educación, la creatividad, la simulación o la comunicación estratégica, pero siempre supeditadas a condiciones estrictas de regulación, supervisión y etiquetado claro. De forma minoritaria, persiste un sector que rechaza cualquier posible legitimación de los *deepfakes* debido a su potencial de daño estructural.

La proyección futura del fenómeno es percibida de manera casi unánime como expansiva. La mayoría de los participantes considera que en los próximos cinco años los *deepfakes* serán cada vez más frecuentes, más realistas y difíciles de detectar, lo que incrementará la presión sobre los sistemas jurídicos, educativos y mediáticos. Este crecimiento es interpretado simultáneamente como una amenaza sistémica y como un desafío tecnológico inevitable.

Adicionalmente la aportación de los expertos refuerza las tendencias detectadas en el alumnado, pero desde una posición aún más precautoria y estructural. Los docentes subrayan el riesgo epistémico que los *deepfakes* suponen para la construcción social de la verdad, la necesidad urgente de alfabetización mediática avanzada, y la importancia de avanzar hacia modelos de gobernanza ética de la inteligencia artificial, que no dependan únicamente de la autorregulación de las plataformas.

Dado el carácter no probabilístico de la muestra, el análisis comparativo debe interpretarse con cautela. En este sentido, las diferencias observadas se presentan como tendencias exploratorias y no como conclusiones generalizables, permitiendo identificar patrones de percepción sin pretensión de extrapolación estadística. La integración de los cuatro ejes analíticos: nivel formativo, ciudad, grado y carácter público o privado de la institución, permite identificar patrones en la construcción de las percepciones sobre los *deepfakes* y el uso de Grok, así como posibles diferencias vinculadas al perfil académico y al contexto institucional.

En primer lugar, el análisis por niveles formativos sugiere una evolución en las percepciones. Los resultados apuntan a una mayor tendencia crítica en el alumnado de 2º curso (Periodismo, US), con una preocupación prioritaria por la desinformación, la manipulación informativa y la erosión de la credibilidad mediática. En 1.º curso (Marketing, UMA) se observa una actitud más abierta hacia la innovación regulada, mientras que en 4.º curso (Publicidad EUSA y Eco-ADE UMA) parece consolidarse una posición intermedia, más reflexiva y orientada a la responsabilidad profesional y al impacto social de la tecnología.

Este patrón se ve reforzado por la variable territorial e institucional. En el caso de Sevilla (US y EUSA), se observa una tendencia hacia una lectura más crítica del fenómeno, aunque con matices según el centro. En la Universidad de Sevilla (pública), especialmente en Periodismo, los *deepfakes* tienden a interpretarse como una amenaza para la democracia informativa, la verificación de fuentes y la confianza pública. En el Centro Universitario EUSA (privado), aunque la preocupación también es elevada, el foco parece desplazarse hacia la imagen, la persuasión publicitaria y los riesgos reputacionales. En Málaga (UMA, pública), por su parte, se aprecia una mayor disposición a explorar usos positivos de la IA generativa en contextos creativos, empresariales y formativos, siempre condicionados a la existencia de marcos regulatorios claros. El análisis por grados refuerza esta lectura. Los datos sugieren una mayor tendencia crítica en el estudiantado de Periodismo, situando la veracidad informativa, la ética pública y la protección frente a la desinformación en el centro de su valoración.

El grado de Publicidad ocupa una posición intermedia, reconociendo el potencial creativo de los *deepfakes*, pero mostrando sensibilidad hacia la suplantación de identidad y los riesgos reputacionales. El alumnado de Marketing parece más proclive a integrar los *deepfakes* en lógicas de innovación y comunicación estratégica,

siempre bajo condiciones de regulación, transparencia y consentimiento. Por su parte, Economía y ADE adopta una perspectiva más pragmática, centrada en riesgos de fraude, ciberdelincuencia y gobernanza digital.

Estas diferencias se ven igualmente moduladas por el carácter público o privado de la institución. En las universidades públicas (US y UMA) se observa una mayor orientación hacia la protección de la veracidad y la responsabilidad social. En la US, esta tendencia parece más acentuada, con una mayor demanda de regulación restrictiva, mientras que en la UMA se aprecia un discurso más orientado al equilibrio entre control normativo e innovación. En la universidad privada (EUSA), el análisis se construye desde una lógica más profesionalizante, vinculada a la ética de la imagen, la reputación de marca y la protección del consumidor.

En lo relativo a Grok como facilitador de *deepfakes*, los datos muestran una valoración mayoritariamente crítica en todos los contextos analizados, aunque con matices. En Periodismo (US) predomina una percepción negativa asociada a la desinformación, mientras que en Publicidad (EUSA) se enfatizan los riesgos para marcas y figuras públicas. En Marketing (UMA) aparece una postura más condicionada, y en Economía y ADE se vincula principalmente a riesgos financieros y de seguridad digital. Estos resultados sugieren que la valoración de la herramienta depende del marco profesional desde el que se interpreta su impacto.

En cuanto a la responsabilidad y la regulación, se observa un consenso transversal en atribuir la responsabilidad principal a quien crea y difunde el contenido, aunque en contextos como Periodismo se introduce con mayor intensidad la idea de corresponsabilidad de plataformas y desarrolladores. La demanda de regulación es generalizada, si bien con matices según el contexto institucional.

Respecto a la proyección futura, los resultados apuntan a un aumento significativo del uso de *deepfakes* en los próximos años. No obstante, se identifican diferencias en la interpretación de sus consecuencias según el ámbito formativo, oscilando entre una percepción de riesgo sistémico y una visión más orientada a oportunidades reguladas. Finalmente, la comparación con el profesorado refuerza estas tendencias desde una perspectiva más estructural, destacando la importancia de la alfabetización mediática, la transparencia algorítmica y la gobernanza ética de la inteligencia artificial.

En su conjunto, la integración de los resultados revela que los *deepfakes* y el uso de Grok constituyen un fenómeno percibido como altamente disruptivo, cuya valoración oscila entre la preocupación ética, la amenaza informativa, el riesgo económico y la oportunidad creativa, en función del nivel formativo, la disciplina, la ciudad y el carácter público o privado de la institución. Esta complejidad refuerza la necesidad de enfoques regulatorios, educativos y profesionales diferenciados, capaces de adaptarse a los distintos contextos desde los que se experimenta el impacto de la inteligencia artificial generativa.

A pesar de que este estudio es de carácter exploratorio, parece que no se distancia mucho de las acciones que se están llevando a cabo en el marco legislativo en España. Así, el Gobierno de España aprobó a principios de enero de 2026, en Consejo de Ministros, el anteproyecto de ley del derecho al honor y en ella incluyó a los *deepfakes* como delitos contra el honor cuando se haga un uso y difusión de imágenes o voces manipuladas con inteligencia artificial sin consentimiento. Un hecho del que se hacían eco los medios de comunicación españoles (Castro, 2026; Menéndez, 2026; Pascual y Sosa Troya, 2026; Piña, 2026).

6. DISCUSIÓN

Los resultados obtenidos permiten, aunque desde una aproximación exploratoria, observar que los *deepfakes* han dejado de ocupar un lugar periférico para integrarse de forma estable en el ecosistema comunicativo actual. Esta transición no se percibe únicamente como un avance tecnológico, sino como un fenómeno con implicaciones directas sobre dimensiones clave como la veracidad informativa, la privacidad o la confianza social. En este sentido, los hallazgos del estudio coinciden con planteamientos recientes que sitúan a los

deepfakes en el centro de la lógica de la posverdad, contribuyendo a la erosión de la credibilidad mediática (Ballesteros-Aguayo y Ruiz del Olmo, 2024). Uno de los aspectos más relevantes es el peso que adquiere la desinformación como principal riesgo identificado por los participantes. Este resultado sugiere que la preocupación no radica tanto en la complejidad técnica de estas herramientas, sino en su capacidad para generar contenidos plausibles que dificultan su detección. Así, la amenaza percibida se vincula más con su impacto comunicativo que con su funcionamiento tecnológico.

Al mismo tiempo, la coexistencia de percepciones negativas (fraude, suplantación de identidad) con valoraciones más positivas (creatividad, uso educativo) pone de manifiesto el carácter híbrido del fenómeno. Esta ambivalencia ya ha sido señalada en la literatura, donde se subraya la dificultad de establecer una definición unívoca de los *deepfakes* debido a la diversidad de sus aplicaciones (Vasist y Krishnan, 2022; Whittaker *et al*., 2023).

Desde el punto de vista ético, los datos reflejan una preocupación transversal en torno a cuestiones como el consentimiento, la privacidad o la veracidad de la información. Este énfasis confirma que los *deepfakes* no se perciben únicamente como una innovación tecnológica, sino como un desafío que afecta directamente a los derechos digitales y a la protección de las personas. En particular, la dificultad para identificar la autoría de los contenidos sintéticos aparece como un elemento que agrava estos riesgos, en línea con investigaciones previas (J. Kaur *et al*., 2024).

En relación con la dimensión normativa, resulta significativo el consenso existente en torno a la necesidad de establecer regulaciones más estrictas. Los participantes no parecen considerar suficientes las medidas actuales basadas en el etiquetado de contenidos, sino que apuntan hacia modelos más exigentes que incluyan mecanismos de trazabilidad, responsabilidad jurídica y protección efectiva de los derechos fundamentales. Esta demanda conecta con los debates recientes sobre la gobernanza de la inteligencia artificial y sus limitaciones actuales.

Por otra parte, la percepción de que los *deepfakes* afectan negativamente a la credibilidad de los medios refuerza la idea de una crisis más amplia en las infraestructuras de autenticidad. No se trata únicamente de contenidos aislados, sino de una transformación estructural en la forma en que se produce y valida la información, tal como ya se ha señalado en la literatura (Gomes-Gonçalves, 2022). En esta misma línea, Westerlund (2019) advierte que uno de los efectos más relevantes de los *deepfakes* es la erosión progresiva de la confianza pública en los contenidos digitales, ya que la creciente dificultad para distinguir entre materiales auténticos y manipulados puede conducir a un escepticismo generalizado hacia la información disponible en el entorno mediático.

Sin embargo, los resultados no son unívocamente negativos, ya que los participantes reconocen también el potencial creativo y educativo de los *deepfakes*, aunque este reconocimiento aparece condicionado. Es decir, su aceptación depende de la existencia de marcos regulatorios claros, supervisión ética y transparencia en su uso. Esta posición refleja una actitud prudente, alejada tanto del rechazo absoluto como de una aceptación acrítica de la tecnología (Kalpokas y Kalpokiene, 2022; Murillo-Ligorred *et al*., 2023).

En lo que respecta a Grok, su valoración mayoritariamente negativa introduce un matiz relevante en el análisis. Los resultados sugieren que herramientas de este tipo, especialmente cuando están integradas en plataformas con alta capacidad de difusión, generan una percepción de riesgo adicional. Esto se relaciona con su potencial para producir contenidos con apariencia de legitimidad, incluso en contextos informativos o académicos, tal como advierten estudios recientes (Acquier y Cossey, 2025; Yamaguchi *et al*., 2025).

Algunos estudios ponen el foco en la regulación para mitigar los efectos negativos de los *deepfakes* y examinan los marcos normativos en diferentes lugares del mundo (Addati, 2025; Jondec Briones *et al*., 2024; Lendvai y Gosztonyi, 2024; Ramos-Zaga, 2024; Santisteban Galarza, 2024), junto con las normas éticas (Fernández Fernández *et al*., 2026). No obstante, desde una perspectiva de gobernanza, los datos apuntan hacia la necesidad de enfoques combinados. La regulación, por sí sola, no parece suficiente si no se acompaña de estrategias educativas orientadas al desarrollo del pensamiento crítico y de tecnologías capaces de detectar contenidos manipulados. En este sentido, la alfabetización mediática emerge como un elemento clave para afrontar los desafíos planteados por los *deepfakes*, en línea con trabajos anteriores (Chapagain *et al*., 2024). Esta necesidad resulta especialmente relevante en el ámbito universitario, donde la formación en competencias críticas frente a los contenidos generados mediante inteligencia artificial se configura como una herramienta esencial para fortalecer la capacidad de evaluación y verificación de la información. En este sentido, Del Moral Pérez *et al*. (2025) destacan que los estudiantes universitarios desarrollan diversas estrategias asociadas al pensamiento crítico para identificar y contrastar contenidos potencialmente falsos generados por IA, subrayando la importancia de reforzar la alfabetización mediática y digital en los entornos de educación superior.

En conjunto, los resultados de este estudio exploratorio permiten afirmar que los *deepfakes* constituyen una tecnología profundamente ambivalente. Si bien ofrecen posibilidades en ámbitos creativos y educativos, también introducen riesgos significativos para la privacidad, la veracidad informativa y la confianza social. Lejos de una aceptación pasiva, la percepción recogida refleja una actitud crítica que pone de manifiesto la necesidad de establecer límites, responsabilidades y marcos de actuación claros en el desarrollo y uso de la inteligencia artificial generativa.

7. CONCLUSIONES

El presente estudio, con la cautela que exige cualquier trabajo de carácter exploratorio, ha permitido examinar cómo el alumnado universitario de tres universidades andaluzas interpreta el fenómeno de los *deepfakes* en el contexto de su integración en herramientas de inteligencia artificial generativa como Grok. Las aportaciones de expertos han contribuido a contextualizar e interpretar los resultados obtenidos desde una perspectiva ética y comunicativa. Más allá de su dimensión técnica, los resultados ponen de relieve que estos contenidos son percibidos como un elemento que incide directamente en cuestiones clave del entorno comunicativo actual, como la veracidad, la identidad o la confianza en la información.

En relación con el nivel de conocimiento y percepción general, los datos muestran que existe una conciencia bastante extendida sobre los riesgos asociados a los *deepfakes*. La desinformación, la suplantación de identidad o los problemas vinculados a la privacidad aparecen de forma recurrente en las respuestas, lo que indica que los participantes no interpretan esta tecnología desde una lógica exclusivamente innovadora, sino también desde una perspectiva preventiva. Aunque se reconoce su potencial en ámbitos creativos o educativos, predomina una actitud prudente.

Las aportaciones de los expertos complementan los resultados obtenidos en el alumnado universitario al enfatizar aspectos relacionados con la opacidad de los sistemas de inteligencia artificial, los efectos sobre la producción de conocimiento y la vulnerabilidad de determinados colectivos. Estas valoraciones permiten contextualizar los riesgos asociados a los *deepfakes* desde una perspectiva ética y comunicativa más amplia.

Respecto a la dimensión regulatoria, los resultados apuntan de forma clara hacia la necesidad de avanzar en marcos normativos más exigentes. Las medidas actuales, centradas en gran medida en el etiquetado de contenidos, son percibidas como insuficientes. En su lugar, se plantea la conveniencia de incorporar mecanismos que permitan identificar el origen de los contenidos, establecer responsabilidades y garantizar una protección efectiva de los derechos de las personas. Esta preocupación conecta con el contexto regulatorio

europeo y con las iniciativas recientes orientadas a limitar los usos no consentidos de tecnologías generativas.

Desde una perspectiva ética, los principios de privacidad, consentimiento, veracidad y transparencia se configuran como ejes fundamentales en la valoración de los *deepfakes*. Los resultados sugieren que la respuesta a este fenómeno no puede depender únicamente de soluciones técnicas o jurídicas, sino que requiere también una dimensión formativa. En este sentido, la alfabetización mediática y la formación en inteligencia artificial aparecen como elementos clave para afrontar estos desafíos de manera crítica.

El análisis específico de Grok introduce un elemento adicional de interés. La herramienta es percibida con cierta desconfianza, especialmente en relación con su capacidad para generar contenidos en entornos de alta difusión. La facilidad de uso y la accesibilidad son interpretadas, en este caso, no solo como ventajas, sino también como factores que pueden intensificar los riesgos asociados a la manipulación y la desinformación. Esto sitúa a las plataformas tecnológicas en una posición relevante dentro del debate sobre la responsabilidad en el desarrollo y uso de estas herramientas.

En cuanto a las limitaciones del estudio, es necesario señalar el tamaño reducido del grupo de profesorado y la concentración geográfica de la muestra, aspectos que condicionan la generalización de los resultados, situándose como una investigación de carácter exploratorio. Futuras investigaciones podrían ampliar el alcance del análisis incorporando muestras más diversas, ampliando la representatividad geográfica y analizando universidades de otras comunidades autónomas e incluso de otros contextos internacionales, con el fin de contrastar la estabilidad de los resultados obtenidos y favorecer una mayor capacidad de generalización. Asimismo, resultaría de interés desarrollar estudios longitudinales que permitan observar la evolución de estas percepciones a medida que la tecnología continúa avanzando. Por otro lado, debe señalarse que el instrumento se basó principalmente en preguntas cerradas orientadas a identificar tendencias generales de percepción sobre los *deepfakes* generados mediante inteligencia artificial y que fue diseñado con una finalidad exploratoria y descriptiva. Aunque permite identificar tendencias generales en las percepciones del alumnado sobre los *deepfakes* generados mediante inteligencia artificial, futuras investigaciones podrían ampliar y validar el cuestionario mediante muestras más diversas e incorporar nuevas variables que permitan profundizar en el análisis de los resultados. Finalmente, futuras investigaciones podrían incorporar indicadores más precisos sobre el nivel de conocimiento, uso y experiencia previa de los participantes con herramientas de inteligencia artificial generativa como Grok, con el fin de profundizar en la interpretación de los resultados obtenidos.

En síntesis, los resultados obtenidos apuntan a que los *deepfakes* constituyen un fenómeno que trasciende lo tecnológico y se sitúa en el centro de debates sociales, éticos y educativos. La incorporación de herramientas como Grok en la vida cotidiana plantea retos que obligan a repensar los mecanismos de confianza en la información y el papel de las instituciones, especialmente las universitarias, en la formación de una ciudadanía crítica y consciente en el contexto digital.

8. REFERENCIAS

Abrar, K. T. (24 de septiembre de 2025). Operationalizing “Freedom of Speech, Not Reach”: A comprehensive AI-driven moderation framework using xAI’s Grok models and Community Notes. *Authorea.* https://doi.org/10.22541/au.175873394.42431625/v1

Acquier, A. y Cossey, J. (2025). *Flooding the academic zone with shit? Academic deepfakes and climate action in a post-truth era.* ESCP Business School Impact Paper. SSRN. https://doi.org/10.2139/ssrn.5467450

Addati, F. A. (2025). Privacidad bajo amenaza digital: la respuesta estadounidense a los *Deepfakes* y el vacío legal en Argentina. *Ratio Iuris*, *13*(1), 197-209. https://id.caicyt.gov.ar/ark:/s23470151/thm14at8c

Ali, A., Ghouri, K.F.K., Neseem, H., Soomro, T.R., Mansoor, W. y Momani, A.M. (2022). *Battle of Deep Fakes: Artificial Intelligence Set to Become a Major Threat to the Individual and National Security* [Conference]. 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates. https://doi.org/10.1109/ICCR56254.2022.9995821

Ballesteros-Aguayo, L. y Ruiz del Olmo, F. J. (2024). Vídeos falsos y desinformación ante la IA: El *deepfake* como vehículo de la posverdad. *Revista de Ciencias de la Comunicación e Información*, 29, 1-14. https://doi.org/10.35742/rcci.2024.29.e294

Bañuelos Capistrán, J. (2020). *Deepfake*: La imagen en tiempos de la posverdad. *Revista Panamericana de Comunicación*, *1*(2), 51-61. https://doi.org/10.21555/rpc.v0i1.2315

Bañuelos Capistrán, J. (2022). Evolución del *deepfake*: Campos semánticos y géneros discursivos (2017-2021). *Icono 14, 20*(2), 267-295. https://doi.org/10.7195/ri14.v20i1.1773

Bustos Díaz, J. y Martin-Vicario, L. (2024). Alfabetización mediática en un mundo hiperconectado: De las redes sociales a la Inteligencia Artificial. *European Public & Social Innovation Review*, 9, 01-17. https://doi.org/10.31637/epsir-2024-1241

Caballero Mariscal, D., Segura, A. y Pinto, M. (2025). El reto de la inteligencia artificial: Un estudio de caso sobre las percepciones del uso del ChatGPT en el estudiantado de Educación e Información y Documentación. *Ibersid, 19*(2), 117-130. https://doi.org/10.54886/ibersid.v19i2.5039

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism, 3*(3), 416–431. https://doi.org/10.1080/21670811.2014.976412

Carpenter, P. (2024). *FAIK: A practical guide to living in a world of deepfakes, disinformation, and AI-generated deceptions.* Wiley.

Castro, I. (12 de enero de 2026). El Gobierno incluye los '*deepfakes*' en el catálogo de intromisiones contra el honor. *Eldiario.es* https://eldiario.es/politica/gobierno-incluye-*deepfakes*-catalogo-delitos-honor_1_12902091.html

Cejudo, R. (2024). Ethical problems of the use of *deepfakes* in the arts and culture. In *Ethics of Artificial Intelligence*. Springer. https://doi.org/10.1007/978-3-031-48135-2_7

Chapagain, D., Kshetri, N. y Aryal, B. (2024). *Deepfake* disasters: A comprehensive review of technology, ethical concerns, countermeasures, and societal implications. En *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)* (pp. 1-9). IEEE. https://doi.org/10.1109/ETNCC63262.2024.10767452

Chesney, R. y Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753-1820. https://scholarship.law.bu.edu/faculty_scholarship/640

Dayal, G., Verma, P. y Sehgal, S. (2024). A comprehensive review on the integration of artificial intelligence in the field of education. *Leveraging AI and Emotional Intelligence in Contemporary Business Organizations*, 331-349. *IG Gklobal Scientific Publishing.* https://doi.org/10.4018/979-8-3693-1902-4.ch020

Del Moral Pérez, M. E., López-Bouzas, N., Castañeda Fernández, J. y Bellver Moreno, M. C. (2025). Universitarios frente a las fake news generadas por la inteligencia artificial: Estrategias asociadas al pensamiento crítico que adoptan. *Educación XX1, 28*(2), 69-121. https://doi.org/10.5944/educxx1.39811

Fernández Fernández, Á., Martínez de Bartolomé Rincón, I. y Morgado Aguirre, B. (2026). Ethics, Personal Image, and Disinformation in the Era of *Deepfakes*. *VISUAL REVIEW. International Visual Culture Review Revista Internacional De Cultura Visual*, *18*(2), 31-46. https://doi.org/10.62161/revvisual.v18.5930

Gregory, S. (2022). *Deepfakes*, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. *Journalism, 23*(3), 521-540. https://doi.org/10.1177/14648849211060644

Gomes-Gonçalves, S. (2022). Los *deepfakes* como una nueva forma de desinformación corporativa: Una revisión de la literatura. *IROCAMM – International Review of Communication and Marketing Mix, 5*(2), 57–74. https://doi.org/10.12795/IROCAMM.2022.v05.i02.02

Guevara Chávez, J. L. y Almeida Macias, M. R. (2025). Infodemia y alfabetización mediática frente a los riesgos de la desinformación automatizada en entornos digitales. *Scripta Mundi*, 4(2), 10-36. https://doi.org/10.53591/scmu.v4i2.2309

Gupta, D. y Fatunmbi, T. O. (2024). Generative AI and *deepfakes*: Ethical implications and detection techniques. *Journal of Science, Technology and Engineering Research*. https://doi.org/10.64206/21rgkc40

Hasani, A., Rasheed, J., Alsubai, S. y Luma-Osmani, S. (2024). Personal rights and intellectual properties in the upcoming era: The rise of *deepfake* technologies. En *Forthcoming Networks and Sustainability in the AIoT Era.* (pp. 379-391). Springer. https://doi.org/10.1007/978-3-031-62881-8_32

Iannuzzi, M., Bogaard, G. y Schell-Leugers, J. M. (2025). AI-generated media and the erosion of trust in legal settings: A review on impact and distortion of belief. *AI & Society*. https://doi.org/10.31234/osf.io/qshbu_v1

Jondec Briones, H. P., Zevallos Loyaga, M. E., Segura Grados, A. Y. y Castrejon Vilchez, E. M. (2024). El uso ilícito de las técnicas de inteligencia artificial y la necesidad de su regulación: el *deepfake*. *Vniversitas Jurídica*, 73. https://doi.org/10.11144/Javeriana.vj73.uiti

Kalpokas, I. y Kalpokiene, J. (2022). *Deepfakes*: A realistic assessment of potentials, risks, and policy regulation. Springer. https://doi.org/10.1007/978-3-030-93802-4

Kamachee, M., Casper, S., Ding, M. L., Yew, R. J. y Cols. (2025). *Video deepfake abuse: How company choices predictably shape misuse patterns*. SSRN. https://doi.org/10.2139/ssrn.5076297

Kaur, A., Noori Hoshyar, A., Wang, X. y Xia, F. (2024). Beyond deception: Exploiting *deepfake* technology for ethical innovation in healthcare. En *Proceedings of the 1st International Conference on Artificial Intelligence in Medicine*. ACM.

Kaur, J., Sharma, K. y Singh, M. P. (2024). Exploring the depth: Ethical considerations, privacy concerns, and security measures in the era of *deepfakes*. En *Navigating the World of Deepfake Technology* (pp. 141-165). IGI Global. https://doi.org/10.4018/979-8-3693-5298-4.ch008

Kılıç, B. y Kahraman, M. E. (2023). Current usage areas of *deepfake* applications with artificial intelligence technology. *İletişim ve Toplum Araştırmaları Dergisi, 6*(2), 45-67. *The Journal of Communication and Social Studies.* https://doi.org/10.59534/jcss.1358318

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D. y Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences, 117*(14), 7684- 7689. https://doi.org/10.1073/pnas.1915768117

Lendvai, G. F. y Gosztonyi, G. (2024). *Deepfake* y desinformación –¿Qué puede hacer el derecho frente a las noticias falsas creadas por *deepfake*? *IDP. Revista de Internet, Derecho y Política*, 41, 1-13, https://doi.org/10.7238/idp.v0i41.427515

Levesque, J. (2025). The AI-enhanced trafficking threat: Examining the technological evolution of trafficking in persons operations with AI tools. *Journal of Human Trafficking*. https://doi.org/10.1080/23322705.2025.2572911

Maniyal, V. y Kumar, V. (2024). Unveiling the *deepfake* dilemma: Framework, classification, and future trajectories. *IT Professional, 26*(3), 45-53. https://doi.org/10.1109/MITP.2024.3369948

Menéndez, M. (13 de enero de 2026). El Gobierno incluye los '*deepfakes*' en los delitos contra el honor y regula los 'true crimes'. https://www.rtve.es/noticias/20260113/gobierno-incluye-*deepfakes*-catalogo-delitos-contra-honor/16891904.shtml

Murillo Ligorred, V. y Ramos Vallecillo, N. (2025). La categorización de los *deepfakes* mediante una investigación cualitativa para la formación de maestros: de la crítica a la creatividad en la inteligencia artificial. *Communiars. Revista De Imagen, Artes Y Educación Crítica Y Social, 8*(14), 45-61. https://dx.doi.org/10.12795/Communiars.2025.i14.03

Murillo-Ligorred, V., Ramos-Vallecillo, N. y Covaleda, I. (2023). Knowledge, integration and scope of *deepfakes* in arts education: The development of critical thinking in postgraduate students. *Education Sciences, 13*(12). https://doi.org/10.3390/educsci13121234

Nannaware, S. C., Pillai, R. y Kate, N. (2025). *Deepfakes* in action: Exploring use cases across industries. In *Deepfakes and Their Impact on Society*. IGI Global. https://doi.org/10.4018/979-8-3693-6890-9.ch004

Okhrymovych, V. (2024). *Deepfake* in the communicative space of Internet culture: Features of application. *Artistic Culture. Topical Issues, 40*(1), 112-125. Artistic Culture. Topical issues. https://doi.org/10.31500/1992-5514.20(2).2024.318261

Pascual, M.G. y Sosa Troya, M. (13 de enero de 2026). El Gobierno busca combatir los '*deepfakes*' incluyéndolos como vulneración al derecho al honor. *El País.* https://elpais.com/sociedad/2026-01-13/el-gobierno-busca-combatir-los-*deepfakes*-incluyendolos-como-vulneracion-al-derecho-al-honor.html

Pedersen, K. T., Pepke, L., Stærmose, T., Papaioannou, M., Choudhary, G. y Dragoni, N. (2025). *Deepfake*-driven social engineering: Threats, detection techniques, and defensive strategies in corporate environments. *Journal of Cybersecurity and Privacy, 5*(2), 18. https://doi.org/10.3390/jcp5020018

Peña-Fernández, S., Meso-Ayerdi, K., Larrondo-Ureta, A. y Díaz-Noci, J. (2023). Without journalists, there is no journalism: The social dimension of generative artificial intelligence in the media. *Profesional de la Información, 32*(2), e320227. https://doi.org/10.3145/epi.2023.mar.27

Piña, R. (12 de enero de 2026). El Gobierno regula para incluir los '*deepfakes*' como un delito contra el honor. *El Mundo*. https://amp.elmundo.es/espana/2026/01/12/69656fc1fdddfff3198b4593.html

Porlezza, C. y Schapals, A. K. (2024). AI Ethics in Journalism (Studies): An Evolving Field Between Research and Practice. *Emerging Media*, *2*(3), 356-370. https://doi.org/10.1177/27523543241288818

Ramos-Zaga, F. (2024). *Deepfake*: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial. Derecho global. *Estudios sobre derecho y justicia*, *9*(27), 359-387. https://doi.org/10.32870/dgedj.v9i27.754

Roe, J., Perkins, M., Somoray, K., Miller, D. y Furze, L. (2025). Can synthetic avatars replace lecturers? An exploratory international study of higher education stakeholder perceptions. *International Journal of Educational Technology in High Education, 22*(71), 1-23. https://doi.org/10.1186/s41239-025-00568-4

Sapkota, R., Raza, S., Shoman, M., Paudel, A. y Karkee, M. (2025). Multimodal large language models for image, text, and speech data augmentation: A survey. *arXiv preprint arXiv, 2*. https://doi.org/10.48550/arXiv.2501.18648

Singh, A. (2024). Development of personality rights and data privacy regulation of *deepfakes*. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.5348028

Sundar, S. S., Kang, H. y Saha, K. (2021). Trust in artificial intelligence: Meta-analysis of user perceptions across domains. *Computers in Human Behavior,* 123, 106878. https://doi.org/10.1016/j.chb.2021.106878

Santisteban Galarza, M. (2024). La criminalización de las ultrafalsificaciones (con especial atención a las implicaciones de la normativa europea de servicios digitales e inteligencia artificial). *Revista de Derecho Penal y Criminología*, 31(ENERO). https://doi.org/10.5944/rdpc.ENERO.2024.38890

Torres-Toukoumidis, A., Marín-Gutiérrez, I. y Pallo-Chiguano, M. (2025). Inteligencia Artificial y alfabetización mediática: análisis bibliométrico (2013-2025). *Cuadernos Del Audiovisual | CAA,* 14. https://doi.org/10.62269/cavcaa.58

Vaccari, C. y Chadwick, A. (2020). *Deepfakes* and disinformation: Exploring the impact of synthetic political media on democratic processes. *Social Media + Society,* *6*(1), 1-13. https://doi.org/10.1177/2056305120903408

Vasist, P. N. y Krishnan, S. (2022). *Deepfakes*: An integrative review of the literature and an agenda for future research. *Communications of the Association for Information Systems,* 51, 1436-1473. https://doi.org/10.17705/1CAIS.05126

Westerlund, M. (2019). The emergence of *deepfake* technology: A review. *Technology Innovation Management Review, 9*(11), 39-52. https://doi.org/10.22215/timreview/1282

Whittaker, L., Mulcahy, R., Letheren, K., Kietzmann, J. y Pitt, L. (2023). Mapping the *deepfake* landscape for innovation: A multidisciplinary systematic review and future research agenda. *Technovation,* 126. https://doi.org/10.1016/j.technovation.2023.102735

Yamaguchi, Y., Hashimoto, C. y Uchino, H. (2025). Fact-checking generative AI in the classroom: Wikipedia, Grok, and the redefinition of critical AI literacy in Japanese higher education based on a 2023 Instructional Practice. En T. Bastiaens (Ed.) P*roceedings of EdMedia + Innovate Learning* (pp. 526-532). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/226182/

CONTRIBUCIONES DE AUTORES/AS, FINANCIACIÓN Y AGRADECIMIENTOS

Contribuciones de los/as autores/as:

Conceptualización: Sánchez-Hunt, Marta; Cestino González, Estefanía y Román-San-Miguel, Aránzazu. **Software:** Sánchez-Hunt, Marta; Cestino González, Estefanía y Román-San-Miguel, Aránzazu. **Validación:** Román San Miguel, Aránzazu. **Análisis formal:** Sánchez-Hunt, Marta; Cestino González, Estefanía y Román-San-Miguel, Aránzazu. **Curación de datos:** Sánchez Hunt, Marta. **Redacción-Preparación del borrador original:** Sánchez-Hunt, Marta; Cestino González, Estefanía y Román-San-Miguel, Aránzazu. **Redacción-Re- visión y Edición:** Sánchez-Hunt, Marta; Cestino González, Estefanía y Román-San-Miguel, Aránzazu. **Visualización:** Sánchez-Hunt, Marta; Cestino González, Estefanía. **Supervisión:** Román San Miguel, Aránzazu. **Administración de proyectos:** No aplica. **Todos los/as autores/as han leído y aceptado la versión publicada del manuscrito:** Sánchez-Hunt, Marta; Cestino González, Estefanía y Román-San-Miguel, Aránzazu.

Financiación: Esta investigación no recibió financiamiento externo.

Conflicto de intereses: Ninguno.

AUTOR/A/ES/AS:
Marta Sánchez-Hunt
Universidad de Cádiz
Profesora en la Universidad de Cádiz, donde imparte docencia en los grados de Publicidad y Relaciones Públicas en materias relacionadas con la comunicación digital y los medios interactivos. Doctora en Comunicación, su labor docente e investigadora se centra en la comunicación política digital, el análisis de redes sociales, especialmente Instagram, la inteligencia artificial y su impacto en los modelos comunicativos contemporáneos. Ha desarrollado investigaciones sobre estrategias de comunicación en redes sociales, desinformación, deepfakes y uso de la inteligencia artificial en el ámbito mediático. Actualmente, una de sus principales líneas de investigación aborda la relación entre inteligencia artificial, ética y comunicación en entornos digitales.
martahunt@uca.es
Índice H: 1
Orcid ID: https://orcid.org/0000-0001-7321-5107
Google Scholar: https://scholar.google.com/citations?user=dCv8USgAAAAJ&hl=es
ResearchGate: https://www.researchgate.net/profile/Marta-Hunt
Academia.edu: https://circulodelestrecho.academia.edu/MartaS%C3%A1nchezHunt

Estefanía Cestino González
Universidad de Málaga
Doctora en Comunicación Audiovisual por el Programa de Doctorado Interuniversitario en Comunicación por la Universidad de Málaga. Licenciada en Comunicación Audiovisual. Maestra de Audición y Lenguaje. Diplomada en Logopedia. Máster en Creación Audiovisual y Artes Escénicas por la Universidad de Málaga y Máster en Neurologopedia Universidad de VIC. Docente en la Universidad de Málaga desde el curso

2010/11. Adscrita al Departamento de Comunicación Audiovisual y Publicidad. Actualmente trabaja en el Departamento de Organización y Administración de Empresas. Pertenece al Proyecto PID2022-138281NB-C22 Repertorios y prácticas mediáticas en la adolescencia y la juventud: recepción y uso de plataformas audiovisuales online.
ecestino@uma.es
Índice H: 7
Orcid ID: https://orcid.org/0000-0002-2436-8665
Google Scholar: https://scholar.google.com/citations?user=pN-GF2oAAAAJ&hl=es
ResearchGate: https://www.researchgate.net/profile/Estefania-Cestino-Gonzalez?ev=hdr_xprf
Academia.edu: https://uma.academia.edu/EstefaniaCestinoGonzalez

Aránzazu Román-San-Miguel
Universidad de Sevilla
Profesora Titular de Periodismo en la Universidad de Sevilla y directora del Máster en Periodismo de Datos e Inteligencia Artificial. Responsable del grupo de investigación Estrategias de Comunicación. Docente en el grado en Periodismo y Doble Grado en Periodismo y Comunicación Audiovisual, donde imparte asignaturas relacionadas con las tecnologías de la información y la comunicación. También es docente en el Máster Universitario en Comunicación Institucional y Política donde coordina la asignatura de Diagnóstico e Investigación de Estrategias de Comunicación Institucional y Política. Una de sus líneas principales de investigación aborda el uso de la Inteligencia Artificial en el ámbito del periodismo y sus límites éticos.
arantxa@us.es
Índice H: 11
Orcid ID: https://orcid.org/0000-0002-9131-2629
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=56041639800
Google Scholar: https://scholar.google.com/citations?user=KQIS_z0AAAAJ&hl=es
ResearchGate: https://www.researchgate.net/profile/Aranzazu-Roman-San-Miguel
Academia.edu: https://independent.academia.edu/aranzazuromansanmiguel

Revista Latina de Comunicación Social

Artículos relacionados:

Ballesteros-Aguayo, L. y Ruiz del Olmo, F. J. (2024). Vídeos falsos y desinformación ante la IA: el *deepfake* como vehículo de la posverdad. *Revista de Ciencias de la Comunicación e Información*, 29, 1–14. https://doi.org/10.35742/rcci.2024.29.e294

Baloğlu, E. y Budak, E. (2025). Cómo integrar la inteligencia artificial en las prácticas periodísticas. *Vivat Academia*, 158, 1–21. https://doi.org/10.15178/va.2025.158.e1604

Bustos Díaz, J. y Martin-Vicario, L. (2024). Alfabetización mediática en un mundo hiperconectado: de las redes sociales a la Inteligencia Artificial. *European Public & Social Innovation Review*, 9, 1–17. https://doi.org/10.31637/epsir-2024-1241

Dan, V. (2026). Deepfakes as a democratic threat: Experimental evidence shows noxious effects that are reducible through journalistic fact checks. *The International Journal of Press/Politics*, *31*(3), 525-550. https://journals.sagepub.com/doi/full/10.1177/19401612251317766

Serrano, J. Y. M., De la Toba Noriega, C. E., Zatarain, S. C., Contreras, L. Y. A., & Lucas, G. Á. D. (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