VISUAL REVIEW | Vol. 18, No. 2, 2026 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.62161/revvisual.v18.6137

DEEPFAKES Y *FACT-CHECKING*
Influencia de la IA en el Ecosistema Desinformativo desde la Perspectiva de Verificadores y Académicos

DAVID GARCÍA-MARÍN [1]
david.garciam@urjc.es
AMAYA NOAIN-SÁNCHEZ [1]
amaya.noain@urjc.es
GUIOMAR SALVAT-MARTINREY [1]
guiomar.salvat@urjc.es
[1] Universidad Rey Juan Carlos, España

PALABRAS CLAVE
Inteligencia artificial
Deepfakes
Desinformación
Fact-checking
Fake news
Periodismo
Verificación informativa

RESUMEN
El objetivo de esta investigación es analizar los usos e impacto de la IA (sobre todo, las deepfakes) en las campañas desinformativas y la utilidad y potencial de esta tecnología en la lucha contra la desinformación. Se realizaron 16 entrevistas semiestructuradas con fact-checkers y expertos en IA y desinformación en el contexto español, analizadas mediante teoría fundamentada. Se destaca cómo las deepfakes asumen un papel tanto simbólico como operativo en el ecosistema desinformativo Su influencia no se limita a la generación de desinformación, sino que afecta también a la propia percepción social de la realidad a través de estrategias como el marketing de falsa bandera y de fenómenos como el dividendo del mentiroso o la enshittification.

Received: 10/ 11 / 2025
Accepted: 19/ 01 / 2026

1. Introducción y marco teórico

1.1. Inteligencia artificial e imágenes desinformativas: las deepfakes

En su configuración actual, el fenómeno de la desinformación saltó a la opinión pública de la mano del término *fake news*, un concepto polémico, aunque rápidamente popularizado y aceptado (Wardle, 2019, 2017). Fue en 2016, durante la campaña del *Brexit* en el Reino Unido y en las elecciones presidenciales celebradas en Estados Unidos cuando representantes políticos usaron dicho término para criticar a los medios de comunicación que no les eran afines (Bastos & Mercea, 2019; Quandt et al., 2019). Todo ello, en un escenario de crisis de credibilidad del periodismo, en el que los medios informativos tradicionales ya habían perdido parte de su influencia y su papel predominante como *gatekeepers* (Allcott & Gentzkow, 2017), a la vez que proliferaban en las redes sociales contenidos manipulados que aparentaban ser informaciones periodísticas, acrecentando el clima de confusión (Tandoc et al., 2018). Desde entonces, el ecosistema desinformativo no ha parado de crecer, particularmente, a raíz de la pandemia de la Covid-19 (Salaverría et al., 2020), cuando se produce un salto cuantitativo en un contexto cada vez más complejo marcado por la sobreabundancia de información o "infoxicación" (WHO, 2020) propulsada, además, en los últimos años por la popularización de la inteligencia artificial (en adelante, IA).

Además de cuantitativamente, la desinformación también ha evolucionado en términos cualitativos. Los avances en la denominada inteligencia artificial generativa han posibilitado el desarrollo de algoritmos capaces de producir contenidos falsos con apariencia de noticias (Flores-Vivar, 2020) y de contenidos audiovisuales veraces. Su implementación en plataformas digitales para atraer a los usuarios ha contribuido a la rápida diseminación de bulos y narrativas engañosas (Bontridder & Poullet, 2021). Entre estos contenidos, el uso más llamativo es el destinado a crear *deepfakes* o contenidos hiperrealistas (Chesney & Citron, 2019), consistentes en imágenes, bien fotografías o vídeos (también audios, pero en menor porcentaje), donde aparecen personas (especialmente figuras relevantes) "ejecutando acciones o emitiendo discursos que jamás realizaron ni pronunciaron" (García-Marín, 2025, p. 79). Estas imitaciones pueden tener usos inocuos como, por ejemplo, el caso de un vídeo de *Buzzfeed* que mostraba al expresidente estadounidense Barack Obama diciendo palabras ofensivas y que se viralizó rápidamente (Vaccari & Chadwick, 2020). Aunque inicialmente el propósito de estos vídeos no sea desinformativo, pueden inducir a error o engaño cuando se consumen de forma descontextualizada, dificultando su correcta decodificación por parte de las audiencias.

Efectivamente, estas "fábricas algorítmicas" (García-Marín, 2025) de producción de imágenes, aunque no se utilicen con intención desinformativa, generan contenidos que pueden llevar a confusión por diferentes motivos. En primer lugar, tanto los datos de entrenamiento como el diseño de los algoritmos pueden contener sesgos que derivan en contenidos discriminatorios u ofensivos hacia determinados grupos sociales. Estudios como el de García-Ull y Melero-Lázaro (2023) demuestran que los sistemas de inteligencia artificial generativa no solo replican los estereotipos de género, sino que los refuerzan e incrementan. En segundo lugar, parte del contenido utilizado por estos sistemas puede estar protegido por derechos de autor, lo que plantea riesgos legales y éticos relacionados con la apropiación indebida de la propiedad intelectual. Finalmente, estas tecnologías utilizan datos circunscritos a un momento temporal específico, lo que limita su capacidad para ofrecer contenidos actualizados y pertinentes en contextos dinámicos.

En todo caso, la proliferación de *deepfakes,* sea o no con intenciones desinformativas, desafía la composición de un ecosistema informativo y una esfera pública saludables. La alteración del contenido audiovisual mediante este tipo de tecnologías abre la puerta a nuevas formas de manipulación ciudadana (Hancock & Bailenson, 2021; Köbis et al., 2021). Su potencial dañino se hizo evidente en los primeros años de popularización de estas herramientas, cuando se utilizaron de manera masiva para generar imágenes falsas de mujeres desnudas (Hao, 2020) y colocarlas en vídeos pornográficos (Cook, 2019). Posteriormente, su uso vinculado a contenidos políticos, por ejemplo, en campañas masivas de desinformación dirigidas a los votantes antes de las elecciones (Dobber et al., 2021), supuso un salto cualitativo (Vaccari & Chadwick, 2020). Sin duda, su capacidad para mermar aún más la confianza en las instituciones públicas y en los propios medios de comunicación preocupa en tanto que puede erosionar los pilares de la democracia (Kalpokas & Kalpokiene, 2022; Wahl-Jorgensen, & Carlson, 2021).

La principal característica de estas producciones es su carácter hiperrealista, lo que favorece su credibilidad entre las audiencias con menor nivel de alfabetización mediática e informacional. Estudios

recientes que abordan el impacto de las *deepfakes* –realizados en diferentes contextos y mediante métodos cuasi-experimentales– muestran que, aun cuando los ciudadanos creen que son capaces de detectar los vídeos manipulados, su capacidad de discernir entre el contenido falso y el verdadero no resulta demasiado elevada (García-Marín et al., 2025; Köbis, et al., 2021). El poder persuasivo de lo audiovisual, mucho mayor que el de los bulos en formato textual, combinado con su rápida difusión, legitima su contenido haciendo que la desinformación tenga mayor apariencia de verosimilitud (Tahir et al., 2021). Además, el hecho de que las herramientas de IA para generar *deepfakes* estén disponibles en línea y sean fácilmente accesibles ayuda a su proliferación.

Las investigaciones iniciales sobre desinformación generada mediante *deepfakes* revelan que las narrativas más frecuentes se centran en la manipulación de la imagen pública de figuras relevantes, especialmente actores políticos (García-Marín, 2025). Asimismo, se ha documentado un uso intensivo de estas tecnologías en contextos bélicos, donde se generan imágenes falsas para demonizar al enemigo, justificar acciones propias y elevar la moral, en línea con estrategias propagandísticas históricas. Otro ámbito preocupante es la creación de este tipo de contenidos (frecuentemente vinculados a campañas extranjeras) destinados a desestabilizar a colectivos concretos, comunidades o grupos sociales en determinados países. Estas producciones falsas elaboradas con IA incluyen desde fotografías donde se muestran falsos ataques protagonizados por migrantes hasta vídeos con recomendaciones médicas fraudulentas, con el objetivo de sembrar el caos, fomentar el odio y erosionar la confianza institucional. Asimismo, se han identificado motivaciones económicas en la generación de este tipo de desinformación, como la promoción engañosa de productos o fraudes financieros a través de la falsa recomendación de tales productos y servicios por parte de figuras relevantes e influyentes. Finalmente, la accesibilidad de las herramientas de elaboración de *deepfakes* ha facilitado la proliferación de contenidos espectaculares sin aparente carga ideológica, orientados a maximizar la viralidad y el beneficio económico en las redes sociales, especialmente entre profesionales del ámbito visual.

1.2. El papel de los fact-checkers contra las deepfakes

Paralelamente al aumento y sofisticación de la desinformación, se ha producido una evolución en las plataformas de *fact-checking*, entidades surgidas en los 2000 en Estados Unidos para comprobar la veracidad de las declaraciones de los actores políticos (Graves, 2016) y que, en la actualidad, verifican todo tipo de contenidos desinformativos. Detectan desde textos, imágenes, audios y vídeos realizados con técnicas tradicionales de edición de video (*cheapfakes*) hasta contenidos más elaborados, generados mediante técnicas algorítmicas como son las *deepfakes* (Maslej et al., 2023). Estas plataformas pueden constituirse como entidades independientes o bien estar vinculadas a medios de comunicación (Graves et al., 2020). En cualquiera de los dos casos, su labor está marcada por la colaboración de los periodistas con diversos perfiles del ámbito tecnológico, tales como expertos en inteligencia de fuentes abiertas (OSINT) (Gregory, 2021) o desarrolladores de *software* y especialistas en análisis forense digital (Ciampaglia, 2018), entre otros, lo que repercute en una gran capacidad de adaptación a los cambios tecnológicos.

Para los *fact-checkers*, la integración de herramientas impulsadas por IA representa un recurso estratégico para hacer frente a la proliferación de desinformación y, en concreto, para detectar *deepfakes* (Brandtzaeg et al., 2018). Desde la aplicación de algoritmos entrenados con técnicas de aprendizaje automático *(machine learning)* y aprendizaje profundo (*deep learning)* para detectar automáticamente el contenido desinformativo, hasta los *chatbots* que permiten interactuar con los usuarios que encuentran contenidos dudosos (Arias-Jiménez et al., 2023), la IA promete ser de gran utilidad en labores de verificación y para evitar la propagación de la desinformación (Pasquetto et al., 2022). Si a esto se le añade el uso de técnicas de análisis forense visual (Gregory, 2021), su labor podría resultar fundamental, por ejemplo, en procesos electorales, coyunturas de crisis como la desencadenada durante la pandemia por Covid-19 (Luengo & García-Marín, 2020) o en procesos de alcance geopolítico como la invasión de Ucrania por Rusia (O'Connor, 2022). A pesar de ello, aún permanecen algunas incógnitas por resolver sobre la efectividad real de los instrumentos de IA en la detección automática de contenidos falsos y en la implementación práctica de estas herramientas en las diferentes fases del proceso de verificación.

Si bien los estudios preliminares elaborados en contextos muy específicos muestran que el impacto de las *deepfakes* en las rutinas de verificación de los *fact-checkers* parece limitado y, por el momento, resultan más difíciles de comprobar las manipulaciones visuales como las descontextualizaciones

(Weikmann, et al., 2023), es predecible que la producción de desinformación mediante instrumentos de IA se vaya perfeccionando de forma progresiva. Es por eso por lo que la capacidad de adaptación a los cambios tecnológicos sitúa a los *fact-chekers* como dique de contención clave contra el contenido desinformativo visual impulsado por IA.

2. Objetivos y metodología

En este contexto, el objetivo central de esta investigación es conocer la influencia que tiene la IA, en especial las *deepfakes*, en el ecosistema desinformativo desde el punto de vista de los profesionales del *fact-checking* e investigadores sobre desinformación e inteligencia aritificial. Como sucede en trabajos previos (Gutiérrez-Caneda & Vázquez-Herrero, 2024; Sánchez González et al., 2022), consideramos que estos perfiles resultan informantes clave para analizar este objeto de estudio por encontrarse en la primera línea de batalla en la lucha contra el fenómeno de la desinformación.

En este estudio, el análisis de la influencia de estos sistemas algorítmicos se desglosa en cuatro elementos, que se abordarán por separado, pero de forma integrada: (1) volumen y complejidad del contenido desinformativo generado con IA (*deepfakes*), (2) usos de la IA en las campañas desinformativas, (3) impacto de la IA en el ecosistema desinformativo, y (4) utilidad y potencial de la IA en la lucha contra la desinformación. Teniendo en cuenta estos cuatro aspectos, los objetivos concretos del trabajo se centran en:

O1. Analizar el volumen y complejidad de la desinformación generada con IA (*deepfakes*) desde la perspectiva de los verificadores e investigadores.

O2. Conocer los usos de la IA en las campañas desinformativas, de acuerdo con la visión de los verificadores e investigadores.

O3. Evaluar el impacto de la IA en el ecosistema desinformativo, también desde la óptica de los expertos anteriormente mencionados.

O4. Analizar cómo evalúan verificadores y académicos la utilidad de los instrumentos de IA en la lucha contra la desinformación.

Esta investigación, de tipo eminentemente cualitativo, se basa en los principios de la teoría fundamentada, tal como fue propuesta por Strauss y Corbin (2002), adoptando un enfoque inductivo orientado a la generación de categorías y a la comprensión emergente de los datos. La teoría fundamentada no solo pretende conocer las dimensiones y categorías clave del fenómeno estudiado, sino sobre todo establecer relaciones entre ellas. Esta metodología permite un análisis sistemático de las experiencias y percepciones de los participantes sin depender de hipótesis preestablecidas (Salvat-Martinrey et al., 2024), lo que facilita la identificación de patrones en el discurso de los sujetos (Ahumada et al., 2025).

Como marco metodológico se empleó la teoría de las representaciones sociales de Moscovici (1979), que proporciona las bases para interpretar cómo los participantes construyen y comunican significados sobre un fenómeno complejo. Este enfoque permitió explorar cómo los sujetos entrevistados desarrollan conocimientos, valores y creencias sobre la influencia de la IA en la desinformación.

2.1. Instrumento y participantes

Se realizó una entrevista semiestructurada a un total de 16 informantes clave (11 *fact-checkers* pertenecientes a agencias de verificación españolas y 5 investigadores expertos en IA y desinformación con afiliación en universidades españolas) (Tabla 1). La presencia de dos perfiles diferenciados (profesionales del *fact-checking* y académicos) permitió la complementariedad en las visiones sobre el objeto de estudio, aunque obligó a adaptar la guía concreta de preguntas a cada uno de los perfiles. Téngase en cuenta que las entrevistas a los profesionales de la verificación no solo se realizaron a periodistas/redactores sino también a miembros de los departamentos de ingeniería y datos de las agencias de verificación.

Las entrevistas giraron en torno a las cuatro dimensiones que se pretenden analizar, anteriormente mencionadas en los objetivos: (1) volumen y complejidad de la desinformación generada con IA, (2) uso de la IA en las campañas desinformativas, (3) impacto de la IA en el ecosistema desinformativo, y (4) utilidad de la IA en la lucha contra la desinformación. De forma prioritaria, las entrevistas se procuraron realizar de forma presencial y fueron grabadas solo en formato sonoro, si bien por motivos operativos gran parte de las mismas tuvieron que llevarse a cabo en remoto utilizando el software Microsoft Teams,

en este caso con grabación en vídeo y audio. Todas se ejecutaron de forma síncrona. Se realizaron durante los meses de febrero y junio de 2025 y su duración se estableció entre los 30 y los 75 minutos.

Tabla 1. Perfil de los sujetos entrevistados.

Nº entrevista	Perfil	Formato
ENT_01	Profesor universitario e investigador en IA y desinformación. Con experiencia profesional previa como *fact-checker.*	Remoto
ENT_02	Profesor universitario e investigador en IA, desinformación y el impacto social de los algoritmos.	Remoto
ENT_03	Profesora universitaria e investigadora en IA y desinformación. Con experiencia profesional previa como *fact-checker.*	Remoto
ENT_04	Profesor universitario e investigador en IA y desinformación.	Remoto
ENT_05	Periodista (*fact-checker*) en agencia de verificación española.	Remoto
ENT_06	Periodista (*fact-checker*) en agencia de verificación española.	Presencial
ENT_07	Periodista (*fact-checker*) en agencia de verificación española con perfil tecnológico.	Presencial
ENT_08	Periodista (*fact-checker)* en agencia de verificación española.	Presencial
ENT_09	Periodista (*fact-checker*) en agencia de verificación española con perfil tecnológico.	Presencial
ENT_10	Miembro del equipo de ingeniería y datos en agencia de verificación española.	Remoto
ENT_11	Miembro del equipo de ingeniería y datos en agencia de verificación española.	Remoto
ENT_12	Periodista (*fact-checker)* en agencia de verificación española.	Remoto
ENT_13	Profesor universitario e investigador en IA y desinformación. Con experiencia profesional previa como *fact-checker.*	Remoto
ENT_14	Periodista (*fact-checker)* en agencia de verificación española.	Presencial
ENT_15	Periodista (*fact-checker)* en agencia de verificación española.	Presencial
ENT_16	Periodista (*fact-checker)* en agencia de verificación española.	Presencial

Fuente: Elaboración propia, 2025.

2.2. Análisis de datos

Se utilizó un enfoque inductivo utilizando los procedimientos de codificación abierta y axial. El proceso de codificación abierta se inició con la transcripción y lectura de las entrevistas con el propósito de profundizar en la comprensión de las categorías emergentes y avanzar hacia la saturación teórica, en consonancia con el método comparativo constante propio de la teoría fundamentada. Se identificaron unidades de significado (palabras o frases) con relevancia conceptual dentro del discurso, de acuerdo con los objetivos planteados. A estas unidades se les asignaron etiquetas descriptivas que funcionaron como marcadores iniciales del contenido. Posteriormente, dichas etiquetas fueron agrupadas en códigos analíticos con base en similitudes temáticas o conceptuales observadas en los datos. Cada código fue respaldado por citas textuales y analizado en función de su fundamentación, es decir, la frecuencia con la que aparecía en el discurso, lo cual permitió identificar patrones recurrentes en el *corpus* analizado.

En una segunda fase, se procedió a la codificación axial. Una vez definidos los códigos durante la codificación abierta, estos fueron organizados en categorías jerárquicas mediante el uso de una matriz paradigmática (Strauss & Corbin, 2002). Este procedimiento permitió identificar relaciones causales, condicionales y contextuales entre los códigos. Las relaciones establecidas incluyeron vínculos como “es parte de”, “explica”, “se caracteriza por” o “es un”, favoreciendo así la integración conceptual. Dada la complejidad visual de las redes de códigos, en el apartado de resultados de este artículo se presentan versiones resumidas y simplificadas de tales redes. Las originales completas pueden consultarse en: https://figshare.com/s/a7b4218c75be1a905ee4.

Finalmente, se revisaron de forma sistemática las categorías y subcategorías, con el objetivo de identificar un hilo teórico que permitiera dar cuenta de la representación social de cada una de las dimensiones analizadas, que se exponen en el siguiente apartado.

Para la presentación de los resultados, se procedió a la anonimización de los participantes. Por ello, los fragmentos extraídos de las entrevistas se presentan bajo la siguiente fórmula: ENT_XX donde XX es el número de entrevista adjudicado en la Tabla 1.

Para el procesamiento de los datos se utilizó el software específico para investigación cualitativa Atlas.ti v.25. El estudio obtuvo la aprobación del Comité de Ética de la universidad del equipo investigador (número de registro interno: 120120250322025).

3. Resultados

3.1. Volumen y complejidad de la desinformación generada con IA

La desinformación generada con instrumentos de IA, aunque crecientemente sofisticada, no parece aún ser mayoritaria. En cierto modo, los participantes en el estudio relativizan el impacto actual de esta tecnología en la producción de desinformación, al menos en términos cuantitativos. Señalan que "no estamos en el terreno de las imágenes a gran escala ni para crear grandes bulos porque todavía no se ha sofisticado mucho" (ENT_13). Además, reconocen que "al desinformador le resulta igual más económico difundir un vídeo fuera de contexto que hacerlo con IA" (ENT_12), lo que sugiere que, por ahora, las estrategias y tecnologías tradicionales siguen siendo más rentables y efectivas. Como consecuencia, el volumen de contenido desinformativo elaborado con IA que los verificadores tienen que abordar no resulta especialmente elevado, al menos "no de una manera tan masiva como al principio podíamos esperar, ya que al principio había voces más alarmistas" (ENT_07).

En este sentido, los expertos entrevistados coinciden en que la falsedad generada con IA aún no constituye la principal fuente de desinformación. Uno de los expertos entrevistados señala que "según datos del Observatorio Europeo de Medios Digitales (EDMO), en torno al 5 y el 10% de las verificaciones se basan en bulos hechos con IA" (ENT_12), aunque reconoce que "este porcentaje era igual o menos del 1% hace pocos años" (ENT_12), lo que indica una tendencia creciente. Sin embargo, se matiza que "no se está viendo una gran oleada de falsedades generadas con IA ni tampoco una gran elaboración, porque los *cheapfakes* ya funcionan, y por eso no hace falta elaborarlos más" (ENT_01). Las *cheapfakes*, manipulaciones burdas (Gamir-Ríos & Tarullo, 2022) que combinan técnicas simples como la contextualización inadecuada o el retoque con instrumentos de edición de imágenes (Schick, 2020), se han convertido en una estrategia eficaz y accesible, por lo que dejan poco espacio para contenidos desinformativos más sofisticados y complejos en su elaboración: "Sí que hay mucho *cheapfake*, mucho vídeo que cualquiera manipula un poco el movimiento de los labios, le metes una voz sintética clonada […] y de eso sí que hemos visto bastante aumento" (ENT_07). La accesibilidad a las herramientas capaces de generar este tipo de contenidos ha democratizado su uso para fines desinformativos, aunque también ha limitado, por el momento, su complejidad técnica en términos generales.

En paralelo, parece evidente que la desinformación generada con IA se encuentra en una fase de expansión y diversificación, caracterizada por una mejora progresiva en la calidad técnica, una creciente accesibilidad y una integración parcial en las dinámicas de manipulación informativa: "Vemos que se utiliza cada vez más, que el acceso es muy fácil, que se populariza, que hay una serie de tendencias en las que el uso es muy creciente" (ENT_06). Aunque aún no domina el ecosistema de la desinformación, su potencial disruptivo genera cierto recelo. Como concluye un experto: "Yo creo que se nos va a complicar a todos, […] el Veo3 de Google es impresionante ya lo que hace" (ENT_14).

En esta línea, el análisis de las entrevistas con verificadores y expertos constata que la irrupción de los instrumentos algorítmicos para crear desinformación ha introducido nuevas complejidades en el ecosistema informativo contemporáneo. La creciente sofisticación de la desinformación generada con IA no solo plantea desafíos técnicos, sino también epistemológicos y metodológicos para el *fact-checking.* Uno de los retos identificados es la opacidad de los sistemas de IA, descritos como "cajas negras" (ENT_01), aspecto que dificulta la trazabilidad del contenido. Como señala uno de los expertos participantes en el estudio, "es muy difícil verificar de dónde viene esa desinformación, porque no se puede rastrear el origen como lo podrías hacer en otro contexto" (ENT_01). En el mismo sentido, otro aspecto relevante es la dificultad metodológica que implica verificar contenidos generados en su totalidad con IA. Los expertos explican que "el hecho de que sea realizada de cero con una herramienta de IA dificulta poder rastrear el origen en muchos casos" (ENT_08). Esta falta de transparencia compromete uno de los pilares fundamentales de la verificación: la identificación de la fuente original. A diferencia de las manipulaciones tradicionales, que partían de una imagen o vídeo real, los contenidos generados íntegramente por IA carecen de un referente verificable, lo que obliga a replantear las estrategias de análisis.

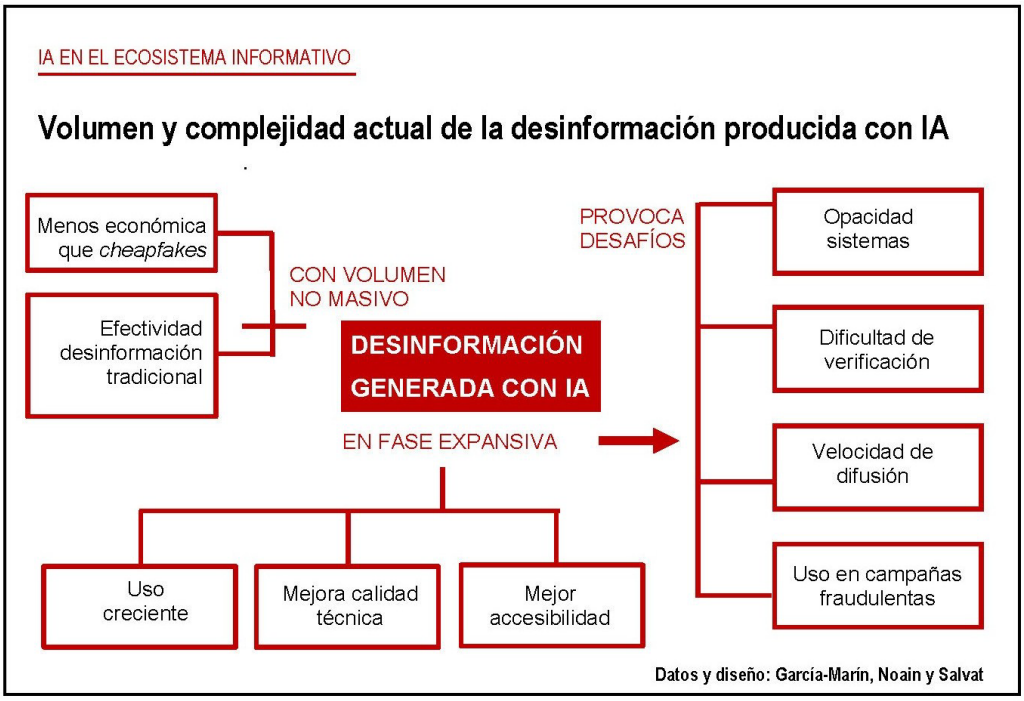
En términos de calidad, también se observa una evolución significativa. Aunque persisten contenidos de baja factura, "hay otros que cada vez tienen más calidad y se están pareciendo más al contenido real" (ENT_03). Esta mejora técnica incrementa el riesgo de suplantación de identidad, especialmente en el caso de figuras públicas. Uno de los expertos advierte: "Lo que más miedo me puede dar son aquellos contenidos que van a suplantar la imagen de una persona relevante y van a poner en su boca palabras

que no ha dicho" (ENT_05). Este tipo de manipulación audiovisual, al explotar la credibilidad de la imagen en movimiento, dificulta la refutación posterior, incluso cuando se dispone de pruebas en contra.

En síntesis, no se trata únicamente de cuánta desinformación se genera con IA, sino de cómo se configura y qué tipo de desafíos presenta su detección y desmentido. Los sujetos entrevistados concluyen que "lo que nos dificulta es la tipología, más que la cantidad" (ENT_07). Esta complejidad se ve agravada por la velocidad de difusión y la facilidad de acceso a herramientas generativas, lo que ha favorecido su uso en campañas fraudulentas, como "la promoción de falsas criptomonedas, plataformas de inversión y falsos medicamentos" (ENT_06).

La Figura 1 recoge la red de códigos sobre el volumen y complejidad de la desinformación producida con IA.

Figura 1. Red de códigos sobre el volumen y complejidad de la desinformación producida con IA.



Fuente: Elaboración propia, 2025.

3.2. Usos de la IA en las campañas desinformativas

En cuanto a la utilización de la IA como herramienta desinformativa (creciente, según lo explicado en el apartado anterior), se perciben dos dimensiones claramente diferenciadas: (1) productiva y (2) estratégica (Figura 2). La dimensión productiva es aquella donde la IA genera de forma directa algún tipo de contenido que pretende ofrecer un mensaje desinformativo o servir como soporte para tales mensajes. Destaca, por tanto, no solo la generación directa de imágenes y audios –la producción de texto con IA se considera residual– sino también el uso de estos sistemas para la elaboración automática de código que permite que los contenidos falsos producidos con IA no sean detectados como, por ejemplo, la elaboración de sistemas para impedir el archivado de páginas fraudulentas. Asimismo, se ha documentado el uso de IA en campañas de *phishing*, generación de *malware*, elaboración de páginas web fraudulentas y anuncios falsos. Estas aplicaciones muestran que la IA no solo afecta al plano simbólico; sino también al operativo, facilitando el fraude y la manipulación a gran escala.

Los usos estratégicos de la IA en las campañas desinformativas resultan menos visibles y complejos de detectar. En primer lugar, se ha documentado la utilización de estos instrumentos para la creación de campañas de desinformación altamente personalizadas. Uno de los expertos explica que "empresas que operan en la *dark web* venden datos de personas con el perfil de personalidad y emocionalidad ligada, lo que permite generar mensajes adaptados a cada individuo" (ENT_04). Estas campañas están siendo gestionadas algorítmicamente con instrumentos de IA, incluyendo la planificación de sus diferentes fases, como la monitorización de reacciones de los usuarios y la evasión de patrones detectables que faciliten su detección: "Cuándo emitir los mensajes desinformativos es algo que se hace

con apoyo de algoritmos, para que los verificadores e investigadores no encuentren un patrón establecido" (ENT_04).

La IA también se utiliza para fomentar la polarización social, por ejemplo, mediante estrategias de *falsa bandera*. El denominado *marketing de falsa bandera* consiste en una táctica comunicativa donde una entidad —ya sea una organización, grupo o individuo— se presenta como otra, usualmente como una parte neutral o incluso antagonista, con el propósito de moldear la opinión pública, inducir determinados comportamientos o desacreditar a un adversario. Este concepto tiene su origen en el ámbito militar y de inteligencia, donde alude a operaciones encubiertas diseñadas para aparentar ser acciones perpetradas por el enemigo. En este tipo de operaciones, los desinformadores utilizan la IA "para producir contenido polémico en redes a fin de crear polaridad y para detectar usuarios en ambos polos" (ENT_04), lo que permite perfilar a los sujetos y dirigir campañas desinformativas adaptadas a cada perfil, gestionadas también con IA, amplificando la fragmentación del espacio público.

Figura 2. Red de códigos sobre los usos de la IA en las campañas desinformativas.

Fuente: Elaboración propia, 2025.

3.3. Impacto de la IA en el ecosistema desinformativo

De acuerdo con la opinión de los participantes en el estudio, rara vez la IA sirve como elemento central de las campañas desinformativas, sino que se utiliza con el fin de reforzar narrativas existentes generando un efecto de arrastre o visibilizador de una campaña lanzada mediante otro tipo de producciones (*cheapfakes*) (Figura 3). En esta labor de respaldo de narrativas desinformativas existentes, los contenidos generados con IA les otorgan "más realismo, más amplitud" (ENT_12), y permiten "sembrar la duda o generar marcos de conversación sobre determinados temas que interesan" (ENT_08).

Uno de los aspectos más inquietantes señalados por los expertos es la capacidad de la IA para generar contenidos altamente persuasivos, incluso cuando son falsos. Como se advierte, "la inteligencia artificial generativa no tiene que ver con la verdad, sino con la retórica y la persuasión, ya que es altamente convincente incluso cuando se equivoca" (ENT_01). Esta característica, sumada al fenómeno de las *alucinaciones* —contenidos fabricados sin base factual—, ha llevado a algunos investigadores a referirse

a la IA como *artificial ignorance*, es decir, una forma de ignorancia automatizada que produce errores con apariencia de veracidad.

El impacto de esta tecnología no se limita a la generación de desinformación, sino que afecta también a la propia percepción social de la realidad. El uso acrítico de la IA puede "empeorar la confianza ciudadana hacia la información" (ENT_01) y fomentar una desconfianza generalizada que lleva a rechazar incluso contenidos verídicos. Para los participantes en el estudio, el efecto acumulativo de estos fenómenos resulta en una creciente sensación de irrealidad: "La gente cada vez se vuelve más descreída, […] ya dudamos incluso de las noticias reales" (ENT_13). Esta situación provoca fenómenos como el denominado *dividendo del mentiroso*, una paradoja comunicativa que surge cuando la proliferación de noticias falsas y campañas de desinformación sistemática erosionan la confianza pública en la veracidad de cualquier contenido, incluso del auténtico. Este fenómeno otorga una ventaja estratégica a quienes mienten deliberadamente, ya que, ante la sospecha generalizada sobre la autenticidad de la información, cualquier sujeto puede negar hechos verdaderos alegando que son falsificaciones o manipulaciones. Este fenómeno permite a los actores deshonestos explotar el escepticismo generado por la desinformación para eludir la rendición de cuentas, deslegitimar pruebas reales o sembrar dudas sobre evidencias verificables.

En la misma línea, uno de los riesgos más señalados es la capacidad de la IA para inundar el espacio informativo con contenidos falsos, desplazando la información veraz. "Se puede crear tan rápido que llegue un punto que, cuando busquemos algo en vías como Internet, aparezca más contenido desinformativo y hecho con IA que real e informativo" (ENT_03). Este fenómeno, conocido como *enshittification*, implica una degradación progresiva del ecosistema digital, donde la sobreabundancia de contenidos generados por IA puede dificultar la distinción entre lo verdadero y lo falso o entre el contenido de calidad y aquel que no la tiene.

Por otro lado, los expertos señalan que la IA no solo actúa como herramienta tecnológica, sino también como agente sociotécnico con capacidad de incidir en los procesos cognitivos, emocionales y culturales de las sociedades contemporáneas. Desde esta perspectiva más estructural, los expertos advierten que los sistemas de IA deben entenderse como agentes cognitivos que no solo procesan información, sino que "influyen en la forma en la que construimos conocimiento, y el modo en que relacionamos experiencias con emociones" (ENT_02). En este sentido, la IA adquiere un poder normativo que trasciende su función instrumental, actuando en ocasiones como "un policía, un juez, un periodista, o un médico" (ENT_04), es decir, como un actor simbólico con capacidad para moldear y transformar la realidad social.

En último término, se advierte sobre un fenómeno emergente: la retroalimentación entre IA y desinformación. Hasta ahora, "la IA se estaba entrenando con contenidos hechos por humanos, pero en adelante va a tomar muchos documentos que ha generado la propia IA" (ENT_15). Este bucle autorreferencial puede consolidar errores, sesgos y falsedades, generando un entorno informativo cada vez más opaco y autorreproducido.

Figura 3. Red de códigos sobre el impacto de la IA como herramienta desinformativa.

Fuente: Elaboración propia, 2025.

3.4. Utilidad de la IA contra la desinformación

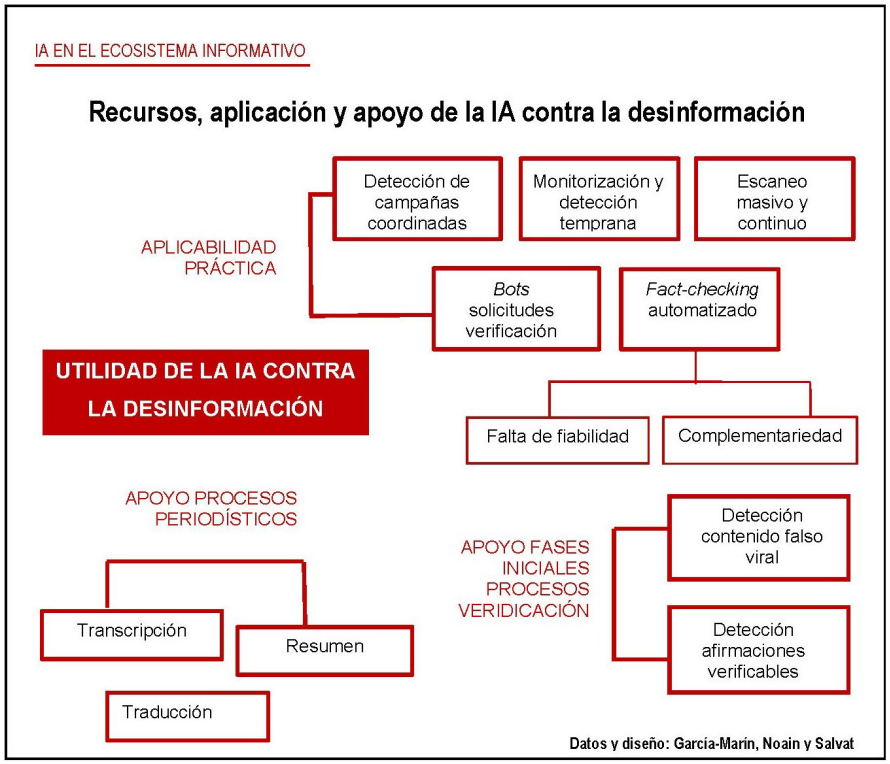
En paralelo al impacto negativo que la IA puede ejercer en el ecosistema informativo actual, resulta evidente el potencial que esta tecnología ofrece para luchar contra las campañas desinformativas. Estos sistemas representan una herramienta valiosa pero limitada en el ecosistema de la verificación, sobre todo en lo que respecta al chequeo automático del contenido potencialmente falso. Su utilidad reside en la automatización de tareas repetitivas (o donde es necesario el procesamiento automático de grandes volúmenes de datos), la ampliación del alcance de la monitorización y la identificación de patrones emergentes. No obstante, su efectividad está condicionada por factores técnicos, epistemológicos y estructurales que, como veremos a continuación, impiden su autonomía plena.

Los sujetos entrevistados coinciden en que estos modelos algorítmicos resultan eficientes en las fases iniciales del proceso de verificación, tales como la detección del contenido falso viral, o de las afirmaciones potencialmente verificables, además de servir de apoyo en procesos que ya están presentes en otras modalidades periodísticas, como el resumen, la traducción o la transcripción de textos (Figura 4). Sin embargo, presentan más dudas sobre su efectividad en la fase de verificación del contenido. De aquí se infiere que, hasta la fecha, la introducción del denominado *fact-checking automatizado* continúa en una fase no exenta de tensiones, limitaciones y desafíos estructurales que condicionan su implementación. Téngase en cuenta que este tipo de verificación consiste en la aplicación de herramientas computacionales para evaluar la veracidad de un contenido potencialmente falso mediante algoritmos que lo analizan en profundidad, lo contrastan con bases de datos, *corpus* documentales o fuentes verificadas, y emiten un juicio sobre su grado de veracidad, todo sin intervención humana directa (García-Marín, 2022). Como se explicará posteriormente, los participantes en el estudio matizan el alcance real de estas tecnologías para la verificación automática.

Como se refería anteriormente, en términos de aplicabilidad práctica, las herramientas de IA han demostrado ser útiles en la identificación de patrones de difusión, análisis de narrativas y detección de campañas coordinadas. Se reconoce un valor instrumental de la IA en las fases preliminares del proceso de verificación, especialmente en la monitorización y detección temprana de contenidos potencialmente desinformativos que circulan en las plataformas digitales. Los verificadores señalan que "en la primera fase de recolección de esos datos o de esas desinformaciones o contenidos engañosos más virales, por ejemplo, la IA nos está empezando a ayudar" (ENT_03). Esta capacidad de escaneo masivo y continuo permite ampliar el alcance de la vigilancia informativa, automatizando tareas que antes requerían una dedicación intensiva de recursos humanos. En este sentido, se destaca que "la IA no se cansa, por lo que la parte de monitorización, la escucha social, la puedes ampliar muchísimo" (ENT_11).

Asimismo, algunas agencias de verificación han desarrollado herramientas específicas para facilitar la identificación de afirmaciones verificables en contenidos audiovisuales. Un ejemplo de ello es el uso de software de vídeo que "ofrece automáticamente las frases verificables de ese contenido audiovisual" (ENT_11), lo cual optimiza el trabajo de los verificadores al reducir el tiempo de análisis preliminar. También se han implementado sistemas de análisis narrativo que permiten detectar "qué narrativas son las más potentes, cómo evolucionan, cuándo suben y cuándo bajan" (ENT_07), lo que resulta útil para anticipar campañas de desinformación estructuradas. Del mismo modo, resulta interesante el uso de "*bots* automáticos para recibir solicitudes de verificación de bulos por parte de la audiencia" (ENT_08), una herramienta que también permite responder automáticamente a tales peticiones si la verificación solicitada ya está disponible en la base de datos del *fact-checker.*

Figura 4. Red de códigos sobre la utilidad de la IA contra la desinformación.



Fuente: Elaboración propia, 2025.

Como se observaba anteriormente, tanto verificadores como expertos coinciden en la limitada efectividad de las herramientas automáticas de detección de falsedades. Aunque existen aplicaciones que ofrecen probabilidades de que un contenido haya sido generado artificialmente, "no hay una herramienta mágica, a día de hoy, que te diga si algo es verdadero o falso, o generado con IA, al 100%" (ENT_12). Estas herramientas, además, presentan resultados inconsistentes, ya que "unas veces aciertan y otras no" (ENT_08), lo que impide su uso como prueba concluyente en el trabajo de verificación. Por ello, se insiste en que "nuestra conclusión de la investigación nunca puede estar basada únicamente en lo que diga la IA" (ENT_06). Además, su efectividad depende de una actualización

constante, ya que "cada día salen 20 herramientas nuevas y no puedes hacer una caja de herramientas y que te valga, ni siquiera, para 6 meses" (ENT_16).

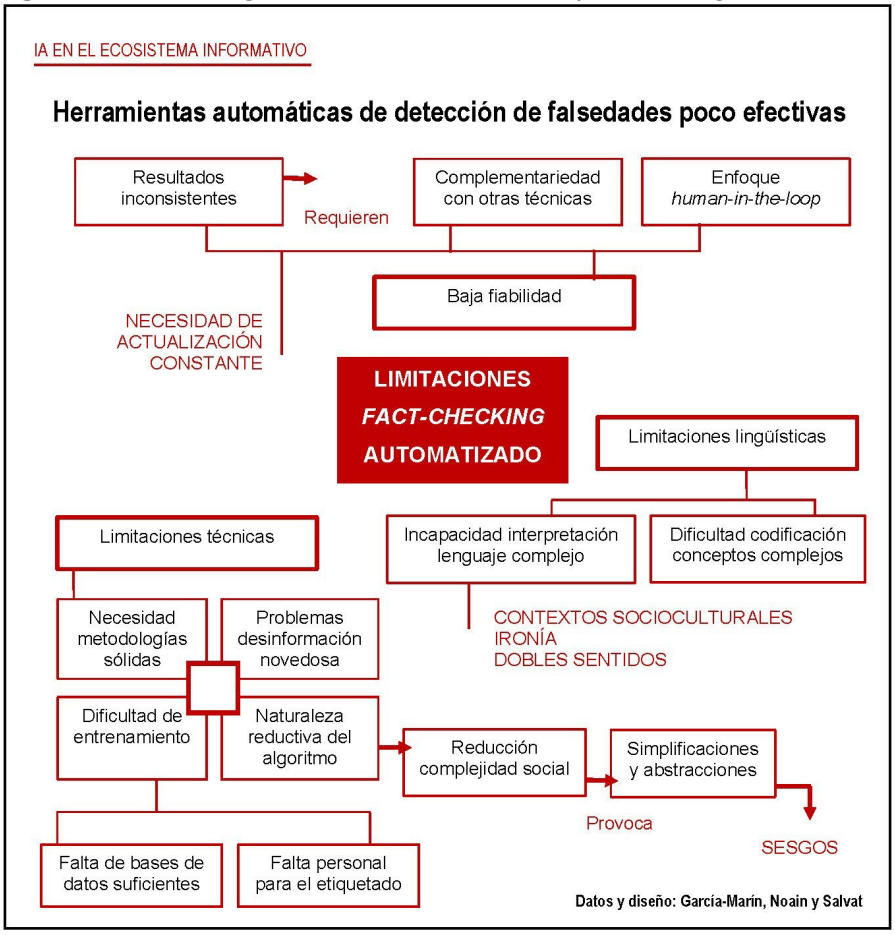
Esta falta de fiabilidad obliga a que el enfoque predominante sea el de *human-in-the-loop*, donde la IA actúa como apoyo, pero nunca como sustituto del análisis periodístico: "Siempre lo primero es el análisis del redactor o los redactores que hemos hecho la investigación, y la herramienta de IA es un apoyo, pero nunca es la base" (ENT_06). Esta complementariedad se justifica también por la incapacidad de estas tecnologías para interpretar elementos complejos del lenguaje como la ironía, los dobles sentidos o los contextos socioculturales. También tienen dificultades para la detección del contenido desinformativo novedoso que no pueda ser comparado con material previo guardado en las bases de datos de los verificadores. Por todo ello, los expertos coinciden en que la IA no puede reemplazar el juicio humano en la fase crítica de la verificación. Como se afirma de forma categórica, "para la verificación, si alguien imagina que hay una herramienta que va a decir de forma automatizada si algo es falso y no se va a necesitar a ningún humano para cerciorarse de ello, está equivocado" (ENT_01). En la misma línea, concluye otro de los entrevistados: "La herramienta te ayuda a llegar quizá antes o mejor, pero siempre es fundamental el conocimiento y la experiencia humana" (ENT_09).

Este papel meramente complementario, siempre en combinación con otras estrategias de verificación, viene determinado por una serie de limitaciones que estos instrumentos manifiestan (Figura 5). En primer lugar, estas herramientas presentan dificultades técnicas derivadas de su entrenamiento. "A veces, no hay suficientes bases de datos para entrenarlas, incluso, tampoco hay suficiente personal" (ENT_01), lo que compromete su fiabilidad y cobertura temática. Expertos y *fact-checkers* subrayan la necesidad de establecer criterios metodológicos sólidos para su entrenamiento, así como una infraestructura de datos adecuada para la elaboración de modelos fiables. "En la verificación no hay un gran *corpus* para poder entrenar estos modelos, ni en inglés ni en español" (ENT_13), lo que limita el desarrollo de sistemas robustos. Lo ideal sería la elaboración de modelos multimodales que integren texto, imagen y contexto, pero esto requiere "mucho etiquetado por parte de humanos para que el entrenamiento sea correcto" (ENT_13).

Además, se identifican problemas estructurales relacionados con el acceso a datos y el control de los algoritmos de las redes sociales donde circula la mayor parte del contenido desinformativo que los verificadores chequean. "No tenemos ese poder mientras que no tengamos el acceso a los algoritmos que están en manos de unas cuantas corporaciones tecnológicas" (ENT_02), lo que genera una dependencia asimétrica de las grandes plataformas. Esta situación se agrava por las restricciones impuestas al uso de APIs, como en el caso de X (antes Twitter): "La última vez que intenté utilizar la API de X para hacer un análisis de discurso de odio, no pude efectuarlo" (ENT_03).

Finalmente, otro obstáculo relevante es la falta de consenso conceptual sobre fenómenos clave que guardan estrecha relación con la desinformación, lo que dificulta su codificación algorítmica. Uno de los expertos entrevistados pone como ejemplo los discursos de odio: "Si no somos capaces de ponernos de acuerdo en qué es un discurso de odio, por ejemplo, va a ser difícil programar una herramienta de detección de este tipo de mensajes" (ENT_02). Esta dificultad se ve amplificada por la naturaleza reductiva de los algoritmos, que "minimizan la complejidad social y hacen simplificaciones y abstracciones" (ENT_11), lo que puede introducir sesgos ideológicos e incorrecciones en los resultados.

Figura 5. Red de códigos sobre las limitaciones del *fact-checking* automatizado.



Fuente: Elaboración propia, 2025.

4. Discusión y conclusiones

La investigación ha permitido analizar el volumen y la complejidad de la desinformación generada con IA (O1); revelando que, aunque este tipo de contenido aún no es mayoritario, su facilidad de uso, evolución técnica y creciente accesibilidad anticipan un aumento significativo en su utilización. Este fenómeno se vuelve especialmente preocupante en contextos de alta polarización o interés estratégico, donde las *deepfakes* podrían adquirir un papel protagónico. A pesar de las suposiciones iniciales que preveían una oleada de uso masivo de este tipo de contenidos sintéticos para desinformar (Chesney & Citron, 2019), los verificadores, expertos y académicos entrevistados en este estudio relativizan su frecuencia actual, en línea con lo descrito por Kalpokas y Kalpokiene (2022). Al menos en términos cuantitativos, el impacto de esta tecnología es aún limitada, aunque cuando se vislumbra una tendencia al alza. Esta baja proliferación aparece fundamentada en dos aspectos: (1) existe un desfase entre el potencial técnico de la IA y su implementación práctica en campañas desinformativas y (2) las *cheapfakes* son más accesibles y rentables, por lo que siguen siendo las manipulaciones más frecuentes (Gamir-Ríos & Tarullo, 2022).

No obstante, el hecho de que cuantitativamente su uso no sea preocupante en la actualidad, no es óbice para menospreciar su futuro potencial desinformativo. La percepción compartida por los entrevistados sugiere que el problema no radica tanto en la cantidad de *deepfakes,* como en el salto cualitativo que supone y en los desafíos metodológicos que plantea su verificación. El hecho de que la IA generativa esté transformando la producción de contenidos falsos, haciéndolos más verosímiles y difíciles de detectar (Bontridder & Poullet, 2021; Flores-Vivar, 2020), anticipa un panorama más complejo para la labor de los verificadores, en un contexto de baja credibilidad en el que las audiencias presentan dificultades para detectar los contenidos falsos elaborados con IA (García-Marín et al., 2025; Köbis et al., 2021).

En relación con los usos de la IA en las campañas desinformativas (O2), los hallazgos muestran que las *deepfakes* operan principalmente como elementos de refuerzo narrativo, más que como núcleo de

las campañas. Esta función amplificadora incrementa su capacidad para sembrar dudas y legitimar discursos falsos, especialmente cuando se integran en estrategias algorítmicas de difusión. Asimismo, la IA se emplea tanto en la generación directa de contenidos como en la planificación estratégica de su circulación, lo que refuerza su papel operativo y simbólico en el ecosistema desinformativo.

Los hallazgos del estudio refuerzan las advertencias de Bontridder y Poullet (2021) sobre el papel de la IA en la personalización de campañas desinformativas y en la planificación algorítmica de su difusión. La dimensión operativa de la IA, evidenciada en su uso para generar *malware*, páginas fraudulentas o campañas de *phishing*, se complementa con una dimensión más sofisticada y menos visible: la gestión algorítmica de campañas altamente personalizadas. Empresas que operan en la *dark web* comercializan perfiles psicológicos y emocionales que permiten adaptar los mensajes desinformativos a cada individuo, optimizando su impacto persuasivo. Esta planificación incluye la monitorización de reacciones y la evasión de patrones detectables, lo que dificulta la labor de los verificadores y desafía los métodos tradicionales de análisis.

Además, la IA se emplea como herramienta para fomentar la polarización social (Torcal & Magalhaes, 2022) mediante estrategias como el *marketing* de falsa bandera (García-Marín, 2025). Estas técnicas permiten a los desinformadores simular identidades antagonistas para moldear la opinión pública y desacreditar adversarios. La IA facilita así la producción de contenido polémico en redes sociales con el objetivo de detectar usuarios en ambos extremos ideológicos, perfilar sus características y dirigir campañas adaptadas a cada grupo.

Nuestro trabajo también subraya que el impacto de la IA (O3) va más allá de la mera producción de contenidos desinformativos. Uno de los retos más relevantes identificados en este estudio es la opacidad de los sistemas de inteligencia artificial, descritos por los participantes como auténticas "cajas negras". Esta característica dificulta de manera significativa la trazabilidad del contenido falso (Morosoli et al., 2025; Wu, 2024), ya que impide rastrear su origen con la precisión que permiten otros contextos. Esta limitación técnica y epistemológica compromete los principios fundamentales del *fact-checking*, especialmente la identificación de fuentes primarias y la reconstrucción del proceso de producción del contenido (Graves, 2016).

La falta de transparencia de los sistemas algorítmicos se vincula con una preocupación más amplia sobre la erosión de la veracidad en el entorno digital, la infoxicación informativa y la denominada *enshittification*, que erosionan la confianza ciudadana y dificultan la distinción entre lo veraz y lo falso. En este sentido, nuestro estudio se alinea con las reflexiones de Kalpokas y Kalpokiene (2022) sobre la anarquía epistémica, en la que los criterios para determinar qué es verdadero se diluyen, generando una suerte de *artificial ignorance* que describe cómo la IA puede generar contenidos erróneos con apariencia de veracidad. La IA afecta directamente la percepción social de la realidad, fomentando una desconfianza generalizada que alcanza no sólo a los contenidos falsos, sino también a los verdaderos. Esta situación refuerza el fenómeno de la decadencia de la verdad descrito por Chesney y Citron (2019), y contribuye al llamado dividendo del mentiroso, en el que la desconfianza generalizada permite a los actores deshonestos negar incluso hechos verídicos. En suma, la retroalimentación entre IA y desinformación, esto es la posibilidad de que los sistemas algorítmicos se entrenen con contenidos generados por la propia IA, plantea riesgos consistentes en la consolidación de errores y sesgos (Deuze & Beckett, 2022), lo que podría agravar la opacidad del entorno informativo y dificultar aún más la labor de verificación.

Finalmente, en relación con el cuarto objetivo, analizar cómo evalúan verificadores y académicos la utilidad de los instrumentos de IA en la lucha contra la desinformación (O4), se concluye que estas herramientas tienen un papel valioso, pero limitado. Esta percepción confirma los resultados de investigaciones anteriores. Para algunos autores, el advenimiento de la IA ha aumentado las posibilidades de combatir la desinformación (Moreno Espinosa et al., 2024; Rubin, 2022) y puede ayudar a distinguir entre información veraz y la distorsión de la realidad (Flores Vivar, 2019; Santos, 2023), incluso reducir el tiempo de detección y aumentar la capacidad de respuesta ante las campañas desinformativas (Cuartielles et al., 2024).

Este potencial de la IA resulta especialmente útil en las fases primigenias, esto es, en tareas de monitorización de contenidos falsos y detección de campañas desinformativas; así como de automatización de procesos. Se pone de manifiesto, por tanto, su aplicación para mejorar la productividad de los *fact-checkers* (Cools & Diakopoulos, 2024; Cools & de Vreese, 2025; Dodds et al., 2025; García de Torres et al., 2025; Wu et al., 2018). No obstante, se subraya que estas herramientas no

pueden sustituir el juicio humano, debido a su limitada fiabilidad, su incapacidad para interpretar contextos complejos y la falta de datos adecuados para su entrenamiento. Esta visión instrumental o complementaria de la IA refuerza el enfoque *human-in-the-loop* (García-Marín, 2022), donde la IA actúa como apoyo, pero no como sustituto del análisis periodístico. Así, aunque se reconoce su utilidad en tareas de monitorización y análisis preliminar, los entrevistados coinciden en que el *fact-checking* automatizado aún presenta limitaciones técnicas y epistemológicas. Esta visión coincide con lo planteado por García-Marín (2022), quien advierte que la verificación automática no puede sustituir el juicio humano, especialmente en contextos complejos.

Asimismo, la implementación efectiva de la IA en el *fact-checking* requiere superar obstáculos técnicos, epistemológicos y estructurales, como la escasez de *corpus* adecuados, el acceso restringido a datos y algoritmos de plataformas digitales, y la falta de consenso conceptual sobre fenómenos clave como los discursos de odio.

4.1. Limitaciones del estudio

Este trabajo presenta tres limitaciones metodológicas. En primer lugar, la muestra se circunscribe al contexto español, lo que limita la extrapolación de los resultados a otros entornos socioculturales y mediáticos. Aunque se ha entrevistado a 16 informantes clave (11 *fact-checkers* y 5 investigadores), todos pertenecen a agencias y universidades españolas. Esto podría implicar una limitación geográfica y cultural que afectaría a la generalización de los resultados a otros contextos internacionales, donde el uso de IA y las dinámicas de desinformación pueden diferir significativamente.

En segundo término, la investigación se basa en entrevistas cualitativas, lo que permite una rica exploración de significados, pero también implica una cierta dependencia de percepciones individuales. Como es habitual en este tipo de trabajos, las opiniones de los participantes podrían estar influenciadas por sus experiencias personales, sesgos profesionales o el momento temporal en que se realizaron las entrevistas.

Por último, el diseño metodológico recurre exclusivamente a entrevistas semiestructuradas, sin incorporar otras técnicas cualitativas (como grupos focales, métodos observacionales o análisis documental) ni métodos cuantitativos que permitan contrastar o complementar los hallazgos. Futuras investigaciones, basadas en estas aproximaciones metodológicas, podrían complementar la profundidad interpretativa de los resultados de nuestro estudio.

5. Agradecimientos

Esta investigación está financiada por el proyecto “Desafíos, usos y limitaciones de la IA en el *fact-checking* y la lucha contra la desinformación” (DESAF_IA) (Ref. 2024/SOLCON-135623) financiado por la convocatoria de Proyectos IMPULSO a la investigación de la Universidad Rey Juan Carlos (2024).

Referencias

Asiri, A., Panday-Shukla, P., Rajeh, H. S., & Yu, Y. (2021). Broadening perspectives on CALL teacher education: From technocentrism to integration. *TESL-EJ, 24*(4).

Ahumada, R., Aravena-Winkler, M., & Sacomori, C. (2025). Social representations of interdisciplinary work from the perspectives of students and clinical tutors on a cardiovascular rehabilitation program in Chile. *International Journal of Educational Research Open*, *9*, 100515. https://doi.org/10.1016/j.ijedro.2025.100515

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Arias Jiménez, B., Rodríguez-Hidalgo, C., Mier-Sanmartín, C., & Coronel-Salas, G. (2023). Use of chatbots for news verification. In P. C. López-López, D. Barredo, Á. Torres-Toukoumidis, A. De-Santis, & Ó. Avilés (Eds.), *Communication and applied technologies* (Vol. 318). Springer. https://doi.org/10.1007/978-981-19-6347-6_12

Bastos, M. T., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review, 37*(1), 38–54. https://doi.org/10.1177/0894439317734157

Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy, 3*(3), 1–21. https://doi.org/10.1017/dap.2021

Brandtzaeg, P. B., Følstad, A., & Chaparro-Domínguez, M.-Á. (2018). How journalists and social media users perceive online fact-checking and verification services. *Journalism Practice, 12*(9), 1109–1129. https://doi.org/10.1080/17512786.2017.1363657

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

Ciampaglia, G. L. (2018). Fighting fake news: A role for computational social science in the fight against digital misinformation. *Journal of Computational Social Science, 1*(1), 147–153. https://doi.org/10.1007/s42001-017-0005-6

Cook, J. (2019, 23 de junio). Here´s what it´s like to see yourself in a deepfake porn video. *The Huffington Post.* https://www. huffingtonpost.co.uk/entry/deepfake-porn-video_uk_5d106e03e4b0aa375f4f1ea7

Cools, H., & de Vreese, C. H. (2025). From Automation to Transformation with AI-Tools: Exploring the Professional Norms and the Perceptions of Responsible AI in a News Organization. *Digital Journalism*, 1–20. https://doi.org/10.1080/21670811.2025.2505982

Cools, H., & Diakopoulos, N. (2024). Uses of Generative AI in the Newsroom: Mapping Journalists' Perceptions of Perils and Possibilities. *Journalism Practice*, 1–19. https://doi.org/10.1080/17512786.2024.2394558

Cuartielles, R., Mauri-Ríos, M. & Rodríguez-Martínez, R. (2024). Transparencia en el uso de la IA en las plataformas de fact-checking en España y sus desafíos éticos. *Communication & Society, 37*(4), 257-271. https://doi.org/10.15581/003.37.4.257-271

Deuze, M., & Beckett, C. (2022). Imagination, Algorithms and News: Developing AI Literacy for Journalism. *Digital Journalism*, *10*(10), 1913–1918. https://doi.org/10.1080/21670811.2022.2119152

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics, 26*(1), 69–91. https://doi.org/10.1177/19401612209443

Dodds, T., Zamith, R., & Lewis, S. C. (2025). The AI turn in journalism: Disruption, adaptation, and democratic futures. *Journalism*. https://doi.org/10.1177/14648849251343518

Flores Vivar, J. M. (2019). Inteligencia artificial y periodismo: diluyendo el impacto de la desinformación y las noticias falsas a través de los bots. *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*, *29*, 197-212. https://doi.org/10.31921/doxacom.n29a10

Flores-Vivar, J. M. (2020). Datos masivos, algoritmización y nuevos medios frente a desinformación y fake news. Bots para minimizar el impacto en las organizaciones. *Comunicación y Hombre, 16*, 101–114. https://doi.org/10.32466/eufv-cyh.2020.16.601.101-114

Gamir-Ríos, J. & Tarullo, R. (2022). Predominio de las *cheapfakes* en redes sociales. Complejidad técnica y funciones textuales de la desinformación desmentida en Argentina durante 2020. *AdComunica*, (23), 97-118. https://doi.org/10.6035/adcomunica.6299

García de Torres, E., Ramos, G., Yezers' ka, L., Gonzales, M., Higuera, L., & Herrera, C. (2025). The use and ethical implications of artificial intelligence, collaboration, and participation in local Ibero-American newsrooms. *Frontiers in Communication*, *10*, 1539844. https://doi.org/10.3389/fcomm.2025.1539844

García-Marín D. (2022). Modelos algorítmicos y fact-checking automatizado. Revisión sistemática de la literatura. *Documentación de las Ciencias de la Información*, *45*(1), 7-16. https://doi.org/10.5209/dcin.77472

García-Marín, D. (2025). *Desinformación y fact-checking en la era de la IA*. Ediciones CIESPAL.

García-Marín, D., Suárez-Álvarez, R., & García-Jiménez, A. (2025). "Todo parece veraz". Credibilidad de la desinformación producida usando IA desde la perspectiva de los estudiantes de comunicación en España. *Revista De Comunicación*, *24*(2), 183–227. https://doi.org/10.26441/RC24.2-2025-3872

García-Ull, F.-J., & Melero-Lázaro, M. (2023). Gender stereotypes in AI-generated images. *Profesional de la información, 32*(5). https://doi.org/10.3145/epi.2023.sep.05

Graves, L. (2016). *Deciding what's true: The rise of political fact-checking in American journalism*. Columbia University Press.

Graves, L., & Lauer, L. (2020). From movement to institution: The "Global Fact" Summit as a field-configuring event. *Sociologica, 14*(2), 157–174. https://doi.org/10.6092/issn.1971-8853/11154

Gregory, S. (2021). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. *Journalism, 23*(3), 708–729. https://doi.org/10.1177/14648849211060644

Gutiérrez-Caneda, B., & Vázquez-Herrero, J. (2024). Redrawing the Lines Against Disinformation: How AI Is Shaping the Present and Future of Fact-checking. *Tripodos*, (55), 55–74. https://doi.org/10.51698/tripodos.2024.55.04

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior and Social Networking, 24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

Hao, K. (2020, October 20). A deepfake bot is being used to "undress" underage girls. *MIT Technology Review*. https://www.technologyreview.com/2020/10/20/1010789/aideepfake-bot-undresses-women-and-underagegirls/

Kalpokas, I., & Kalpokiene, J. (2022). On alarmism: Between infodemic and epistemic anarchy. In I. Kalpokas & J. Kalpokiene (Eds.), *Deepfakes: A realistic assessment of potentials, risks, and policy regulation* (pp. 41–53). Springer.

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience, 24*(11), 103364. https://doi.org/10.1016/j.isci.2021.103364

Luengo, M., & García-Marín, D. (2020). The performance of truth: Politicians, fact-checking journalism, and the struggle to tackle COVID-19 misinformation. *American Journal of Cultural Sociology, 8*(3), 405–427. https://doi.org/10.1057/s41290-020-00115-w

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R. (2023). *The AI Index 2023 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/report

Moreno Espinosa, P., Abdulsalam Alsarayreh, R. A., & Figuereo Benítez, J. C. (2024). El Big Data y la inteligencia artificial como soluciones a la desinformación. *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*, *38*, 437-451. https://doi.org/10.31921/doxacom.n38a2029

Morosoli, S., Resendez, V., Naudts, L., Helberger, N., & de Vreese, C. (2025). "I Resist". A Study of Individual Attitudes Towards Generative AI in Journalism and Acts of Resistance, Risk Perceptions, Trust and Credibility. *Digital Journalism*, 1–20. https://doi.org/10.1080/21670811.2024.2435579

Moscovici, S. (1979). *El psicoanálisis, su imagen y su público*. Huemul.

O'Connor, C. (2022). TikTok is at the forefront of Russia's propaganda war. https://www.isdglobal.org/isd-in-the-news/ciaran-oconnor-tiktok-is-at-the-forefront-of-russias-propaganda-war/

Pasquetto, I. V., Jahani, E., Atreja, S., & Baum, M. (2022). Social debunking of misinformation on WhatsApp: The case for strong and in-group ties. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW1), 1–35. https://doi.org/10.1145/3512964

Quandt, T., Frischlich, L., Boberg, S., & Schatto-Eckrodt, T. (2019). Fake news. *The International Encyclopedia of Journalism Studies*, 1–6. https://doi.org/10.1002/9781118841570.iejs0128

Rubin, V. L. (2022). *Misinformation and disinformation: Detecting fakes with the eye and AI*. Springer.

Salaverría, R., Buslón, N., López-Pan, F., León, B., López-Goñi, I., & Erviti, M.-C. (2020). Desinformación en tiempos de pandemia: Tipología de los bulos sobre la Covid-19. *Profesional de la Información, 29*(3), e290315. https://doi.org/10.3145/epi.2020.may.15

Salvat-Martinrey, G., García-Marín, D., & Zorogastua, J. (2024). La inteligencia artificial generativa y su impacto en el sector de la comunicación. Percepción de los futuros profesionales. *RAE-IC, Revista de la Asociación Española de Investigación de la Comunicación*, *11*, raeic11e06. https://doi.org/10.24137/raeic.11.e.6

Sánchez-González, M., Sánchez-Gonzales, H.M., & Martínez-Gonzalo, S. (2022). Inteligencia artificial en verificadores hispanos de la red IFCN: proyectos innovadores y percepción de expertos y profesionales. *Estudios sobre el Mensaje Periodístico, 28*(4), 867-879. https://dx.doi.org/10.5209/esmp.82735

Santos, F. C. C. (2023). Artificial intelligence in automated detection of disinformation: A thematic analysis. *Journalism and Media*, *4*(2), 679–687. https://doi.org/10.3390/journalmedia4020043

Schick, N. (2020, 22 de diciembre). Don't underestimate the cheapfake. *MIT Technology Review.* https://www.technologyreview.com/2020/12/22/1015442/cheapfakes-more-political-damage-2020-election-than-deepfakes/

Strauss, A. L., & Corbin, J. (2002). *Bases de la investigación cualitativa: Técnicas y procedimientos para desarrollar la teoría fundada.* Editorial Universidad de Antioquia.

Tahir, R., Batool, B., Jamshed, H., Jameel, M., Anwar, M., Ahmed, F., Zaffar, M. A., & Zaffar, M. F. (2021). Seeing is believing: Exploring perceptual differences in deepfake videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

Tandoc Jr., E. C., Lim, Z.-W., & Ling, R. (2018). Defining 'fake news'. *Digital Journalism, 6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Torcal, M., & Magalhaes, P. C. (2022). Ideological extremism, perceived party system polarization, and support for democracy. *European Political Science Review*, *14*(2), 188–205. https://doi.org/10.1017/S1755773922000066

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

Wahl-Jorgensen, K., & Carlson, M. (2021). Conjecturing fearful futures: Journalistic discourses on deepfakes. *Journalism Practice, 15*(6), 803–820. https://doi.org/10.1080/17512786.2021.1908838

Wardle, C. (2019). *First draft's essential guide to understanding information disorder*. First Draft. https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Consejo de Europa. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Weikmann, T., & Lecheler, S. (2023). Cutting through the hype: Understanding the implications of deepfakes for the fact-checking actor-network. *Digital Journalism, 12*(10), 1505–1522. https://doi.org/10.1080/21670811.2023.2194665

WHO. World Health Organization. (2020). *Novel coronavirus (2019-nCoV). Situation report - 13*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf

Wu, S. (2024). Journalists as individual users of artificial intelligence: Examining journalists' "value-motivated use" of ChatGPT and other AI tools within and without the newsroom. *Journalism*. https://doi.org/10.1177/14648849241303047

Wu, S., Tandoc, E. C., & Salmon, C. T. (2018). Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production. *Journalism Studies*, *20*(10), 1440–1457. https://doi.org/10.1080/1461670X.2018.1521299