VISUAL REVIEW | Vol. 18, No. 7, 2026 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.62161/revvisual.v18.6474

INTELIGENCIA ARTIFICIAL Y VERIFICACIÓN INFORMATIVA FRENTE A DESINFORMACIÓN DIGITAL: RENDIMIENTO, CONFIANZA Y ÉTICA

Rendimiento técnico y percepción ciudadana en procesos de *fact-checking* automatizado

ANA LUCÍA DE VEGA MARTÍN [1]
analucia.devega@universidadviu.com
[1]Universidad Internacional de Valencia, España

PALABRAS CLAVE

Inteligencia artificial
Verificación de la información
Desinformación
Redes sociales
Confianza pública
Transparencia
Alfabetización mediática

RESUMEN

La proliferación de la desinformación digital impulsa el uso de sistemas de inteligencia artificial (IA) en la verificación informativa. Este estudio analiza el rendimiento técnico de estas herramientas y la confianza ciudadana en su aplicación al fact-checking. Con diseño mixto, se examinan 150 informes de verificación que integran IA y una encuesta a 600 usuarios de redes sociales de España y América Latina. Los resultados muestran elevada eficacia en la verificación automatizada y una confianza moderada, condicionada por la transparencia algorítmica y los sesgos. Se concluye que los modelos híbridos IA-humanos constituyen la estrategia más sólida para la verificación informativa.

RECIBIDO: 05/ 08 / 2026
ACEPTADO: 20/ 10 / 2026

1. Introducción

En la sociedad digital contemporánea, la circulación masiva de información a través de redes sociales y plataformas digitales ha transformado profundamente la forma en que la ciudadanía accede, consume y evalúa los contenidos informativos. Este escenario, si bien ha ampliado las posibilidades de participación y acceso al conocimiento, ha favorecido también un incremento sostenido de la desinformación, entendida como un fenómeno que compromete la veracidad de los discursos públicos y condiciona la toma de decisiones individuales y colectivas (Wardle & Derakhshan, 2021). Los estudios más recientes sobre el papel de la inteligencia artificial en este contexto destacan los desafíos y enfoques actuales para combatir la desinformación, subrayando no solo las capacidades técnicas de detección, sino también las dificultades relacionadas con sesgos, transparencia y gobernanza de los sistemas automatizados (Saeidnia et al., 2025).

La desinformación se ha consolidado como un problema estructural del ecosistema mediático global, con efectos directos sobre la confianza institucional y la calidad democrática. Episodios recientes, como la difusión de contenidos falsos durante la pandemia de COVID-19, evidenciaron la vulnerabilidad informativa de amplios sectores sociales y la dificultad de distinguir entre información contrastada y contenidos manipulados en entornos digitales (Mian & Khan, 2020). Este contexto ha reforzado la necesidad de fortalecer la alfabetización mediática e informacional, concebida como una competencia crítica para acceder, analizar y producir información de manera ética y responsable (De Vega Martín, 2023).

En este contexto, la verificación informativa adquiere una relevancia que trasciende su dimensión técnica y se inscribe en el terreno de la mediación simbólica y cultural de la verdad en los entornos digitales. La creciente automatización de los procesos de fact-checking no solo modifica los tiempos y las escalas de la verificación, sino que introduce nuevas formas de autoridad informativa, en las que algoritmos y sistemas automatizados participan activamente en la construcción de lo que se considera fiable, verdadero o legítimo. Desde esta perspectiva, la desinformación no constituye únicamente un problema de circulación de contenidos falsos, sino un fenómeno que interpela a los modos contemporáneos de producción, validación y visualización de la información en el ecosistema mediático digital.

1.1. Inteligencia artificial y verificación informativa

En respuesta a estos desafíos, la inteligencia artificial (IA) ha adquirido un papel central en los procesos de verificación informativa. Los avances en aprendizaje automático y procesamiento del lenguaje natural han permitido desarrollar sistemas capaces de identificar patrones discursivos, analizar grandes volúmenes de contenido y evaluar la fiabilidad de la información en tiempo real (Kreps et al., 2022). Su incorporación al ámbito del *fact-checking* ha contribuido a mejorar significativamente la rapidez y la eficiencia en la detección de contenidos falsos o manipulados (Zhou et al., 2022; Kreps & Monin, 2024).

En los últimos años, distintas iniciativas han aplicado soluciones basadas en IA para limitar la propagación de desinformación. En el contexto español destacan proyectos como IVERES, impulsado por RTVE para la detección temprana de contenidos desinformativos (Rigall, 2024), o VerificAudio, desarrollado por PRISA Media en colaboración con Google News Initiative, orientado a la identificación de *deepfakes* sonoros (PRISA Media, 2025). A nivel internacional, grandes plataformas tecnológicas han incorporado modelos automatizados que combinan verificación algorítmica y bases de datos contrastadas (García, Liu & Zellers, 2023).

1.2. Desafíos éticos y sociales de la verificación automatizada

Pese a su potencial técnico, el uso de la IA en la verificación informativa plantea relevantes desafíos éticos y sociales. Diversos estudios advierten que estos sistemas pueden reproducir sesgos derivados de los datos con los que son entrenados, afectando a la neutralidad y equidad de los resultados (Bender et al., 2021). Asimismo, la opacidad algorítmica y la limitada explicabilidad de los procesos automatizados contribuyen a generar desconfianza pública y cuestionan la legitimidad social de estas herramientas (García et al., 2023).

Desde una perspectiva comunicativa y educativa, la incorporación de la IA en la gestión de la información no puede entenderse únicamente como un avance tecnológico. Supone también un reto formativo que exige promover la comprensión crítica de los algoritmos, la ética de los datos y el papel

de la supervisión humana. En este sentido, la alfabetización mediática e informacional se presenta como un eje clave para favorecer un uso responsable de la tecnología y para mitigar los riesgos asociados a la automatización de la verificación informativa.

En este sentido, la incorporación de sistemas automatizados de verificación plantea interrogantes que afectan directamente a la relación entre tecnología, confianza y cultura informativa. La delegación parcial de la verificación en sistemas algorítmicos reconfigura los criterios tradicionales de credibilidad y desplaza, al menos en parte, el juicio humano como instancia exclusiva de validación informativa. Esta transformación resulta especialmente significativa en un entorno mediático caracterizado por la centralidad de lo visual, la circulación acelerada de contenidos multimodales y la creciente opacidad de los procesos tecnológicos. Analizar la verificación automatizada implica, por tanto, examinar no solo su eficacia operativa, sino también su impacto en las formas contemporáneas de comprensión, interpretación y evaluación crítica de la información por parte de la ciudadanía.

1.3. Propósito y objetivos del estudio

A partir de este contexto, el presente estudio tiene como propósito analizar el papel de la inteligencia artificial en la verificación informativa frente a la desinformación digital, atendiendo tanto a su rendimiento técnico como a la percepción y el grado de confianza ciudadana en su aplicación. De manera específica, se plantean los siguientes objetivos:

1. Examinar el rendimiento de los sistemas de verificación informativa basados en inteligencia artificial mediante el análisis de plataformas y proyectos que integran herramientas automatizadas de *fact-checking*.
2. Analizar la percepción y el nivel de confianza de los usuarios de redes sociales en los sistemas de verificación basados en IA, considerando los factores éticos y sociales que condicionan su aceptación.

La relevancia de esta investigación reside en ofrecer una aproximación empírica que articula la eficacia técnica de la IA con su legitimidad social, aportando evidencias sobre la brecha existente entre rendimiento automatizado y confianza ciudadana. Desde esta perspectiva, el estudio contribuye al debate académico sobre la verificación informativa en entornos digitales y subraya la importancia de modelos híbridos, basados en la complementariedad entre automatización y supervisión humana, así como de estrategias de alfabetización mediática orientadas a fortalecer la competencia informacional de la ciudadanía.

El estudio permite situar la verificación informativa basada en inteligencia artificial como un proceso en el que la eficacia técnica y la supervisión humana se articulan en la construcción de credibilidad informativa. La combinación de análisis de rendimiento y percepción ciudadana aporta evidencia empírica sobre la relación entre automatización, confianza y legitimidad de los procesos de verificación en entornos digitales.

2. Diseño y método

La investigación adopta un enfoque metodológico mixto, con un diseño descriptivo y explicativo de carácter no experimental y transversal, adecuado para analizar fenómenos complejos que combinan dimensiones técnicas y sociales, como es el caso de la verificación informativa basada en inteligencia artificial (IA). Este enfoque permite integrar el análisis del rendimiento de sistemas automatizados con el estudio de la percepción y la confianza ciudadana, ofreciendo una visión articulada entre eficacia tecnológica y legitimidad social (Creswell & Creswell, 2018; Silver, 2024).

El objeto formal del estudio es el uso de sistemas de IA en los procesos de verificación informativa frente a la desinformación digital, abordado desde una doble perspectiva analítica. Por un lado, se examina el funcionamiento y desempeño técnico de herramientas de verificación automatizada empleadas por plataformas especializadas. Por otro, se analizan las actitudes, percepciones y niveles de confianza de los usuarios de redes sociales respecto a la aplicación de estas tecnologías en la detección de contenidos falsos o manipulados.

Dado el carácter descriptivo y explicativo del estudio, no se formulan hipótesis causales, optándose por un enfoque basado en objetivos y variables de análisis que permiten explorar patrones, tendencias y relaciones relevantes entre el rendimiento de la IA y su aceptación social. El ámbito de estudio se sitúa

en el ecosistema digital hispanohablante, con especial atención a España y América Latina, tanto en lo relativo a las plataformas de verificación analizadas como a la procedencia de los participantes en la encuesta.

La investigación se organiza en dos fases complementarias (Tabla 1): (i) un análisis documental y de contenido de informes de verificación elaborados por plataformas que integran IA, orientado a caracterizar su rendimiento, métodos y limitaciones; y (ii) una encuesta a usuarios de redes sociales destinada a medir percepción, conocimiento y confianza en la verificación automatizada. La revisión de literatura se emplea como apoyo para la contextualización teórica y la definición de categorías analíticas, no como método independiente, dado el carácter empírico del estudio.

Tabla 1. Enfoque metodológico de las fases

Fase del estudio	Método	Objetivo	Técnicas empleadas
Análisis de plataformas	Cualitativo	Evaluar la implementación de IA en fact-checking	Revisión documental y análisis de contenido
Encuesta a usuarios	Cuantitativo	Medir percepción y confianza en IA aplicada a fact-checking	Cuestionario estructurado con escala Likert

Fuente: Elaboración propia.

Nota: la revisión documental se emplea como técnica de apoyo para la contextualización teórica, no como método independiente de investigación.

2.1. Fase 1: análisis de plataformas de verificación con inteligencia artificial

La primera fase se centra en el análisis documental y de contenido de 150 informes de verificación publicados por plataformas que incorporan herramientas de IA en sus procesos de *fact-checking*. La selección de las plataformas se realizó a partir de criterios de inclusión y exclusión definidos según la literatura especializada y los estándares internacionales de verificación digital (Wardle & Derakhshan, 2021), garantizando la presencia de metodologías verificables de uso de IA, una trayectoria consolidada y un compromiso explícito con la transparencia de los procesos de verificación (Tabla 2).

Tabla 2. Criterios de selección de las plataformas de verificación de noticias

Criterio de selección	Justificación
Uso documentado de IA en fact-checking	Se seleccionaron únicamente plataformas con metodologías verificadas en el uso de IA.
Relevancia en el ecosistema digital	Plataformas acreditadas por la International Fact-Checking Network (IFCN) y otros organismos especializados.
Trayectoria y fiabilidad	Se incluyeron plataformas con al menos tres años de actividad verificadora documentada.
Cobertura geográfica	Se incluyeron plataformas con actividad en España y América Latina para garantizar representatividad en el ámbito hispanohablante.
Acceso a bases de datos verificadas	Garantiza que las herramientas de IA puedan comparar información con fuentes fiables en tiempo real.
Transparencia en los modelos de IA	Se excluyeron plataformas que no explicitan los criterios de decisión de sus algoritmos.

Fuente: Elaboración propia.

Los informes analizados, correspondientes a un periodo temporal de doce meses (noviembre 2024-octubre 2025), constituyen la unidad de análisis de esta fase. El procedimiento analítico se organizó en torno a un sistema de categorías previamente definido, estructurado en tres dimensiones principales (Tabla 3): eficiencia de la IA, métodos empleados y desafíos y limitaciones. Estas dimensiones permitieron evaluar aspectos como la precisión en la detección de desinformación, el tiempo de respuesta de los sistemas automatizados, el uso de técnicas de procesamiento del lenguaje natural y aprendizaje profundo, así como la identificación de sesgos algorítmicos, falsos positivos y mecanismos de supervisión humana.

Tabla 3. Dimensiones e indicadores de análisis de la Fase 1

Dimensión de análisis	Indicadores específicos
Eficiencia de la IA	Precisión en la detección de desinformación; tiempo de respuesta; volumen de contenido procesado
Métodos empleados	Uso de procesamiento del lenguaje natural; redes neuronales; aprendizaje profundo; sistemas híbridos con supervisión humana
Desafíos y limitaciones	Presencia de sesgos algorítmicos; nivel de transparencia metodológica; identificación de falsos positivos y falsos negativos; límites de explicabilidad

Fuente: Elaboración propia.

Este análisis permitió obtener una visión sistemática del grado de automatización de los procesos de verificación y de las principales fortalezas y limitaciones técnicas de los sistemas basados en IA, aportando una base empírica para la interpretación posterior de los resultados de la encuesta.

2.2. Fase 2: encuesta sobre percepción pública de la IA en la verificación informativa

La segunda fase adopta un enfoque cuantitativo y tiene como objetivo analizar la percepción pública y el grado de confianza ciudadana en el uso de la inteligencia artificial (IA) aplicada a la verificación informativa. En coherencia con los objetivos del estudio, esta fase se centra en tres variables principales: eficacia percibida, nivel de conocimiento sobre la IA aplicada al *fact-checking* y grado de confianza en los sistemas de verificación automatizada. Asimismo, se incorporan variables contextuales relacionadas con los hábitos de consumo informativo y las prácticas personales de verificación, consideradas relevantes para interpretar las actitudes hacia la verificación automatizada.

La recogida de datos de la encuesta se desarrolló de forma independiente al análisis documental de informes, como una fase diferenciada orientada a examinar la percepción ciudadana sobre la verificación informativa basada en inteligencia artificial.

2.2.1. Muestra, muestreo y ámbito de estudio

La muestra estuvo compuesta por 600 usuarios de redes sociales, seleccionados mediante muestreo no probabilístico por cuotas, atendiendo a criterios de edad, país de residencia, nivel educativo y experiencia previa con plataformas de *fact-checking* (Tabla 4). Este tipo de muestreo se consideró adecuado ante la inexistencia de un marco muestral universal de usuarios de redes sociales y por su capacidad para garantizar diversidad interna en perfiles sociodemográficos y niveles de alfabetización digital.

Tabla 4. Criterios de selección de la muestra

Criterio	Descripción y justificación
Edad	Se incluyeron participantes entre 18 y 65 años, rango de edad con mayor consumo de información en entornos digitales (Ipsos, 2023).
País de residencia	Se seleccionaron participantes de España, México, Argentina y Colombia, países con verificadores activos acreditados por IFCN y presencia de fenómenos de desinformación relevantes.
Nivel educativo	Se incluyeron individuos con educación secundaria, universitaria y de posgrado, para analizar diferencias en la confianza en IA según nivel académico.
Experiencia con *fact-checking*	Se diferenciaron usuarios que han utilizado plataformas de verificación de información y aquellos que desconocen su existencia.

Fuente: Elaboración propia.

El ámbito territorial del estudio incluyó España, México, Argentina y Colombia, países seleccionados por su relevancia en el ecosistema mediático hispanohablante y la presencia de iniciativas activas de verificación informativa. El rango de edad de los participantes se situó entre los 18 y los 65 años, franja que concentra un mayor consumo informativo en entornos digitales.

2.2.2. Instrumento y estructura del cuestionario

La recogida de datos se realizó mediante un cuestionario estructurado, aplicado de forma remota a través de un formulario digital. El instrumento incluyó 20 ítems, organizados en cuatro bloques temáticos, diseñados para operacionalizar las variables del estudio.

El Bloque 1, dedicado a los datos sociodemográficos, permitió caracterizar el perfil de los participantes en términos de edad, país de residencia, nivel educativo y vinculación con ámbitos de comunicación, periodismo o tecnología (Tabla 5).

Tabla 5. Bloque 1. Datos sociodemográficos

Pregunta	Tipología de respuesta
¿Cuál es su edad?	Opción múltiple: (18-24), (25-34), (35-44), (45-54), (55-65)
¿En qué país reside actualmente?	Opción múltiple: España, México, Argentina, Colombia
¿Cuál es su nivel educativo?	Opción múltiple: Secundaria, Universitario, Posgrado
¿Trabaja o estudia en el ámbito de la comunicación, periodismo o tecnología?	Sí/No

Fuente: Elaboración propia.

El Bloque 2 se centró en los hábitos de consumo informativo y prácticas de verificación, incluyendo la frecuencia de consumo de noticias en redes sociales, los canales de acceso a la información, la experiencia previa con noticias falsas y las estrategias empleadas ante la sospecha de desinformación (Tabla 6).

Tabla 6. Bloque 2. Hábitos de consumo informativo

Pregunta	Tipología de respuesta
¿Con qué frecuencia consume noticias en redes sociales?	Opción múltiple: Varias veces al día, Una vez al día, Varias veces a la semana, Rara vez, Nunca
¿A través de qué canales consume noticias con mayor frecuencia?	Opción múltiple: Redes sociales, Portales web, Televisión, Radio, Prensa escrita
¿Ha detectado alguna vez una noticia falsa en redes sociales?	Sí/No
¿Qué hace cuando sospecha que una noticia puede ser falsa?	Opción múltiple: Busca fuentes oficiales, Consulta verificadores de datos, Comparte la noticia sin verificar, No hace nada

Fuente: Elaboración propia.

El Bloque 3 abordó el conocimiento y la confianza en la IA aplicada al *fact-checking*, mediante preguntas dicotómicas y ítems tipo Likert de cinco puntos, orientados a evaluar la percepción de eficacia de la IA, el nivel de confianza generado por estas herramientas y la valoración de la complementariedad entre sistemas automatizados y verificadores humanos (Tabla 7).

Tabla 7. Bloque 3. Conocimiento y confianza en la IA aplicada al fact.checking

Pregunta	Tipología de respuesta
¿Conoce o ha utilizado alguna plataforma de *fact-checking*?	Sí/No
¿Cree que la IA puede detectar noticias falsas de manera más eficaz que un humano?	Escala Likert (1=Totalmente en desacuerdo, 5=Totalmente de acuerdo)
¿Cuánta confianza le generan las herramientas de IA para verificar noticias?	Escala Likert (1=Sin confianza, 5=Mucha confianza)
¿Cree que los verificadores humanos deberían complementar los sistemas de IA en la detección de *fake news*?	Sí/No

Fuente: Elaboración propia.

Finalmente, el Bloque 4 analizó la percepción de la efectividad y las limitaciones de la IA en la verificación informativa, incluyendo la aceptación de mecanismos automatizados en plataformas digitales, la identificación de limitaciones percibidas (como la falta de transparencia, los sesgos algorítmicos o los problemas técnicos) y las expectativas sobre la posible sustitución de los verificadores humanos (Tabla 8).

Tabla 8. Bloque 4. Percepción de la efectividad de la IA en la verificación de noticias

Pregunta	Tipología de respuesta
¿Cree que las redes sociales deberían incorporar IA para identificar y eliminar noticias falsas automáticamente?	Sí/No
¿Cuáles cree que son las principales limitaciones de la IA en la verificación de noticias?	Opción múltiple: Falta de transparencia, Sesgos algorítmicos, Falta de datos fiables, Problemas técnicos
¿Considera que en el futuro la IA sustituirá por completo a los verificadores humanos?	Sí/No

Fuente: Elaboración propia.

2.2.3. Procedimiento de aplicación y control de calidad

La captación de participantes se realizó mediante una combinación de difusión en redes sociales, envío del cuestionario a bases de datos académicas universitarias y colaboración con organizaciones de *fact-checking*, que facilitaron la distribución del instrumento a través de sus canales de divulgación. La aplicación del cuestionario se desarrolló durante tres meses (entre septiembre y noviembre de 2025), siguiendo una estrategia escalonada de difusión para maximizar la diversidad de perfiles.

Con el fin de asegurar la calidad del instrumento, el cuestionario fue sometido a un proceso de validación por juicio de expertos, en el que cinco especialistas en comunicación digital e inteligencia artificial evaluaron la claridad, pertinencia y coherencia de los ítems. Asimismo, se llevó a cabo una prueba piloto con 20 participantes, que permitió comprobar la comprensión de las preguntas, la consistencia de las respuestas y el tiempo medio de cumplimentación, estimado entre 5 y 7 minutos.

2.2.4. Consideraciones muestrales y éticas

Se observó que una proporción de los participantes presentaba vinculación con ámbitos profesionales relacionados con la comunicación, el periodismo o la tecnología. Este aspecto se consideró relevante para los fines del estudio, dado el papel activo de estos perfiles en la creación, circulación y verificación de contenidos informativos, y se tuvo en cuenta en la interpretación de los resultados.

El estudio se desarrolló conforme a los principios éticos aplicables a investigaciones sociales de carácter no experimental y de riesgo mínimo. La participación fue voluntaria y anónima, sin recogida de datos personales identificativos. Antes de iniciar el cuestionario, los participantes recibieron información clara sobre los objetivos del estudio, el uso exclusivamente académico de los datos y su derecho a abandonar la encuesta en cualquier momento. Las respuestas fueron tratadas de forma agregada, garantizando la confidencialidad y un tratamiento ético de la información.

3. Trabajo de campo y análisis de datos

El trabajo de campo se desarrolló en dos fases empíricas, en coherencia con el diseño metodológico del estudio: el análisis de informes de verificación elaborados por plataformas que integran inteligencia artificial (IA) y la aplicación de una encuesta estructurada a usuarios de redes sociales.

3.1. Fase 1: recopilación y análisis de informes de verificación

La primera fase consistió en el análisis de 150 informes de verificación publicados por plataformas de *fact-checking* que incorporan IA en sus procesos, correspondientes al periodo comprendido entre noviembre de 2024 y octubre de 2025. Cada informe constituyó la unidad de análisis.

Conviene precisar que los valores cuantitativos de rendimiento técnico recogidos en esta fase (como precisión, tiempo de verificación y volumen de contenidos) se basan en las métricas y descripciones reportadas en los propios informes y documentación pública de las plataformas analizadas, por lo que reflejan el desempeño declarado en dichos materiales y no corresponden a una auditoría independiente realizada por el equipo investigador.

Los documentos fueron recopilados y organizados en una base de datos para su tratamiento sistemático. El análisis se llevó a cabo mediante análisis de contenido, aplicando una matriz de categorías previamente definida y alineada con los objetivos del estudio. La codificación permitió registrar indicadores relacionados con la eficiencia técnica (precisión declarada, tiempos de respuesta

y volumen de contenido procesado), los métodos empleados (tipo de tecnología utilizada y grado de intervención humana) y las limitaciones señaladas en los propios informes. El procedimiento combinó el recuento sistemático de variables observables con una descripción analítica de las tecnologías y enfoques de verificación empleados por las plataformas.

3.2. Fase 2: aplicación y tratamiento de la encuesta

La segunda fase se desarrolló mediante un formulario online aplicada a 600 usuarios de redes sociales de España, México, Argentina y Colombia, durante un periodo de tres meses (septiembre–noviembre de 2025). La captación de participantes se realizó a través de canales digitales y redes académicas, con el objetivo de maximizar la diversidad de perfiles sociodemográficos y de experiencias informativas.

Una vez finalizado el trabajo de campo, los datos fueron sometidos a un proceso de depuración previo al análisis. El tratamiento estadístico se realizó mediante estadística descriptiva, utilizando frecuencias y porcentajes para las variables categóricas, y medias y desviaciones típicas para los ítems medidos mediante escalas Likert. Este análisis permitió identificar patrones generales de percepción, conocimiento y confianza en la IA aplicada a la verificación informativa, siguiendo la estructura de los bloques definidos en el cuestionario.

3.3. Consideraciones aplicadas durante el trabajo de campo

Durante el desarrollo del trabajo de campo se aplicaron las condiciones de participación y tratamiento de datos establecidas en el diseño metodológico, garantizando la voluntariedad de la participación y el carácter anónimo de las respuestas, así como el tratamiento agregado de la información recogida.

4. Resultados

4.1. Resultados de la Fase 1: análisis de informes de verificación basados en IA

El análisis de los 150 informes de verificación permitió identificar patrones consistentes en relación con el rendimiento técnico de los sistemas basados en inteligencia artificial, las tecnologías empleadas en los procesos de *fact-checking* y las principales limitaciones operativas señaladas por las propias plataformas.

4.1.1. Precisión, velocidad y capacidad de procesamiento

Los resultados muestran que los sistemas de verificación automatizada alcanzan niveles elevados de precisión en la detección de contenidos desinformativos. Tal como se recoge en la Tabla 9, la precisión media se situó en el 92,8 %, con un rango que oscila entre el 87 % y el 98 %, lo que evidencia diferencias relevantes entre plataformas, aunque sin descensos pronunciados que comprometan la eficacia global del sistema.

En términos de eficiencia temporal, el tiempo medio de verificación por contenido fue de 4,1 segundos, con valores mínimos de 2,5 segundos y máximos de 6,8 segundos. Esta variabilidad indica que, si bien todos los sistemas operan en márgenes de tiempo reducidos, existen diferencias asociadas al tipo de tecnología empleada y al grado de intervención humana requerido en determinadas verificaciones.

Respecto a la capacidad de procesamiento, las plataformas analizadas gestionaron un volumen medio de 11.200 contenidos verificados por día, con registros que alcanzaron hasta 19.700 contenidos diarios en los sistemas con mayor grado de automatización. Estos datos ponen de manifiesto la escalabilidad de los sistemas basados en IA para afrontar grandes volúmenes de información en contextos de alta circulación de contenidos.

Tabla 9. Evaluación de las características téncicas de las plataformas

Indicador evaluado	Valor promedio	Rango observado
Precisión en detección de fake news	92,8%	87%-98%
Tiempo promedio de verificación por noticia	4,1 segundos	2,5 - 6,8 segundos
Volumen de noticias analizadas por día	11.200 artículos	6.500 - 19.700

Fuente: Elaboración propia.

La presentación de los indicadores de precisión, velocidad y volumen de procesamiento permite contextualizar el alcance operativo de los sistemas de verificación automatizada analizados y comparar de forma sistemática los distintos enfoques tecnológicos empleados por las plataformas.

4.1.2. Tecnologías empleadas y niveles de rendimiento asociados

El análisis de los informes permitió identificar una distribución desigual de las tecnologías utilizadas en los procesos de verificación (Tabla 10). El procesamiento del lenguaje natural (PLN) se empleó en el 91 % de las plataformas analizadas, constituyéndose como la tecnología base más extendida. Estas plataformas registraron una eficiencia media del 89,7 %, situándose ligeramente por debajo de otros enfoques más complejos.

Tabla 10. Tecnología utilizada en la detección de desinformación con IA

Método empleado	Porcentaje de plataformas que lo utilizan	Eficiencia promedio (%)
Procesamiento de Lenguaje Natural (PLN)	91%	89,7%
Aprendizaje profundo con redes neuronales	76%	92,4%
Modelos híbridos (IA + verificadores humanos)	69%	95,1%
Análisis de imágenes y *deepfakes*	42%	86,3%

Fuente: Elaboración propia.

La distribución de tecnologías observada pone de manifiesto diferencias relevantes en el grado de automatización y en el rendimiento medio de los sistemas de verificación, ofreciendo una base empírica para analizar la complementariedad entre enfoques automatizados y supervisión humana.

Los modelos de aprendizaje profundo basados en redes neuronales, presentes en el 76 % de las plataformas, alcanzaron una eficiencia media superior (92,4 %), lo que indica un mejor desempeño en la identificación de patrones complejos y relaciones no lineales en los contenidos analizados.

Por su parte, los modelos híbridos, que combinan sistemas automatizados con supervisión humana, estuvieron presentes en el 69 % de las plataformas y mostraron los valores más elevados de eficiencia media (95,1 %). Este grupo presenta, además, una menor dispersión en los resultados, lo que sugiere un comportamiento más estable en los procesos de verificación.

El análisis de imágenes y detección de *deepfakes*, identificado en el 42 % de las plataformas, mostró una eficiencia media del 86,3 %, situándose como el enfoque con menor rendimiento relativo. No obstante, estos sistemas se aplicaron mayoritariamente a contenidos audiovisuales específicos, lo que limita su comparabilidad directa con los métodos centrados en contenido textual.

4.1.3. Necesidad de supervisión humana y limitaciones operativas

El análisis de los informes de verificación evidenció referencias recurrentes a la necesidad de validación humana, especialmente en casos de corrección de falsos positivos y falsos negativos, así como en contenidos con alta carga contextual. Asimismo, los informes describen limitaciones relacionadas con la explicabilidad de los procesos algorítmicos y con la trazabilidad de las decisiones de verificación, tanto en sistemas basados en PLN como en modelos de aprendizaje profundo. En el caso del análisis de contenidos multimodales, se identifican restricciones técnicas en la detección de manipulaciones complejas, especialmente en materiales audiovisuales de baja calidad o con alteraciones sutiles.

Estos resultados ponen de relieve que, incluso en contextos de alto rendimiento automatizado, la verificación informativa continúa dependiendo de procesos de contextualización y validación que no pueden resolverse exclusivamente mediante sistemas algorítmicos.

4.2. Resultados de la Fase 2: percepción ciudadana sobre la IA en la verificación informativa

4.2.1. Características sociodemográficas de la muestra

La Tabla 11 recoge la distribución sociodemográfica de la muestra. En términos de edad, los grupos más representados corresponden a los intervalos 25–34 años (28 %) y 35–44 años (25 %), seguidos del grupo 18–24 años (21 %). Los tramos de mayor edad presentan una presencia menor, con un 16 % en el grupo de 45–54 años y un 10 % en el de 55–65 años.

Tabla 11. Datos sociodemográficos de la muestra

Variable	Categorías	Porcentaje
Edad	18-24 años	21%
	25-34 años	28%
	35-44 años	25%
	45-54 años	16%
	55-65 años	10%
País de referencia	España	35%
	México	30%
	Argentina	20%
	Colombia	15%
Nivel Educativo	Secundaria	35%
	Universitario	52%
	Posgrado	13%
Vinculación con el ámbito de la comunicación, periodismo o tecnología.	Sí	27%
	No	73%

Fuente: Elaboración propia.

En cuanto al país de residencia, la participación se concentra en España (35 %) y México (30 %), mientras que Argentina (20 %) y Colombia (15 %) presentan porcentajes inferiores, coherentes con los criterios de captación establecidos. Respecto al nivel educativo, más de la mitad de la muestra cuenta con estudios universitarios (52 %) y un 13 % con estudios de posgrado. Finalmente, el 27 % de los participantes declara vinculación profesional o académica con los ámbitos de la comunicación, el periodismo o la tecnología.

4.2.2. Hábitos de consumo informativo en redes sociales

La Tabla 12 muestra una alta frecuencia de consumo informativo en redes sociales. El 58 % de los participantes declara consultar noticias varias veces al día y el 24 % al menos una vez al día, frente a un 18 % que lo hace con menor frecuencia. En relación con la exposición a desinformación, el 72 % afirma identificar noticias falsas de manera diaria o varias veces por semana. No obstante, un 38 % reconoce haber compartido noticias sin verificar. Entre estos participantes, el motivo más frecuente fue que el contenido coincidía con sus creencias (62 %), seguido de la percepción de fiabilidad de la fuente (21 %) y de la falta de tiempo para verificar (11 %).

Tabla 12. Hábitos de consumo informativo

Variable	Categorías	Porcentaje
Frecuencia de consumo de noticias en redes sociales	Varias veces al día	58%
	Una vez al día	24%
	Varias veces a la semana	12%
	Rara vez	6%
Frecuencia con la que identifica noticias falsas	Todos los días	26%
	Varias veces a la semana	46%
	Ocasionalmente	20%
	Nunca	8%
Ha compartido noticias sin verificar	Sí	38%
	No	62%

Motivo para compartir noticias sin verificar	Coincidía con sus creencias	62%
	Parecía confiable	21%
	No tenía tiempo para verificar	11%
	No consideró que fuera relevante	6%

Fuente: Elaboración propia.

Los valores medios y la dispersión de estas variables se recogen en la Tabla 13. La frecuencia de consumo informativo presenta una media elevada (M = 3,34; DT = 0,87), mientras que la frecuencia de identificación de noticias falsas se sitúa en un valor inferior (M = 2,90; DT = 0,89). La variable relativa a compartir noticias sin verificar muestra una desviación típica baja (DT = 0,49), lo que indica una distribución relativamente homogénea de esta conducta en la muestra.

Tabla 13. Medias y DT Bloque 2

Variable	Media	Desviación Típica (DT)
Frecuencia de consumo de noticias en redes sociales (1-4)	3.34	.87
Frecuencia de identificación de noticias falsas (1-4)	2.90	.89
Ha compartido noticias sin verificar (1=No, 2=Sí)	1.38	.49
Motivo principal para compartir noticias sin verificar (1-4)	1.61	.78

Fuente: Elaboración propia.

4.2.3. Conocimiento y confianza en la IA aplicada al fact-checking

La Tabla 14 presenta los resultados relativos al nivel de conocimiento y confianza en la IA aplicada a la verificación informativa. El 61 % de los encuestados declara no conocer el uso de IA en procesos de *fact-checking*, mientras que un 18 % afirma haber utilizado herramientas basadas en IA para verificar noticias. El resto se sitúa en posiciones intermedias de familiaridad. En términos de confianza, el 39 % manifiesta niveles altos, el 35 % niveles moderados y el 26 % niveles bajos. Entre los motivos de desconfianza, la falta de transparencia en los procesos es el más señalado (51 %), seguido de los posibles sesgos algorítmicos (28 %).

Tabla 14. Conocimiento y confianza en la IA aplicada al fact-checking

Variable	Categorías	Porcentaje
Nivel de conocimiento sobre IA en fact-checking	No conoce el uso de IA en fact-checking	61%
	Ha oído hablar del tema, pero no sabe cómo funciona	21%
	Conoce la IA en fact-checking, pero no la ha utilizado	15%
	Ha utilizado herramientas basadas en IA para verificar noticias	18%
Nivel de confianza en la IA aplicada a fact-checking	Alta confianza (4-5 en escala Likert)	39%
	Confianza moderada (3 en escala Likert)	35%
	Baja confianza (1-2 en escala Likert)	26%
Motivos de desconfianza en la IA aplicada a fact-checking	Falta de transparencia en los procesos	51%
	Posibles sesgos algorítmicos	28%
	No confía en ninguna verificación automatizada	21%

Fuente: Elaboración propia.

Los valores medios y la dispersión de estas variables se detallan en la Tabla 15. El nivel de conocimiento presenta una media baja (M = 1,75; DT = 0,96), mientras que la confianza en la IA se sitúa en un valor intermedio (M = 3,07; DT = 1,11), con una desviación típica relativamente elevada que refleja heterogeneidad en las respuestas.

Tabla 15. Medias y DT Bloque 3

Variable	Media	Desviación Típica (DT)
Nivel de conocimiento sobre IA en *fact-checking* (1-4)	1.75	.96
Nivel de confianza en IA aplicada a *fact-checking* (1-5)	3.07	1.11
Motivos de desconfianza en IA (1=No desconfía, 2=Desconfía)	1.51	.50

Fuente: Elaboración propia.

4.2.4. Percepción de la efectividad y limitaciones de la IA en la verificación informativa

La Tabla 16 recoge la valoración de la efectividad de la IA en comparación con la verificación humana. El 43 % de los participantes considera que la IA es más efectiva, mientras que un 29 % señala que los verificadores humanos son indispensables y un 28 % opta por un modelo híbrido como alternativa preferente. En relación con las limitaciones percibidas, la falta de transparencia en la toma de decisiones es la opción más señalada (47 %), seguida de los posibles sesgos algorítmicos (32 %) y de la dificultad para interpretar ironía, sarcasmo o contexto cultural (21 %).

Tabla 16. Percepción de la efectividad de la IA en la verificación de noticias

Variable	Categorías	Porcentaje
Percepción sobre la efectividad de la IA en fact-checking	Más efectiva que los humanos	43%
	Los verificadores humanos son indispensables	29%
	Modelo híbrido (IA + humanos) es la mejor opción	28%
Principales limitaciones percibidas en la IA aplicada a fact-checking	Falta de transparencia en la toma de decisiones	47%
	Posibles sesgos algorítmicos	32%
	Incapacidad para interpretar ironía, sarcasmo o contexto cultural	21%
¿Considera que en el futuro la IA sustituirá a los verificadores humanos?	Sí	18%
	No	82%

Fuente: Elaboración propia.

La Tabla 17 muestra las medias y desviaciones típicas asociadas a estas percepciones. La efectividad percibida de la IA presenta una media cercana al valor intermedio de la escala (M = 1,99; DT = 0,81), mientras que las limitaciones percibidas muestran una media moderada (M = 1,47; DT = 0,50). La expectativa de sustitución completa de los verificadores humanos por IA presenta una media baja (M = 1,18; DT = 0,38), lo que indica una distribución concentrada en la opción negativa.

Tabla 17. Medias y DT Bloque 4

Variable	Media	Desviación Típica (DT)
Percepción sobre efectividad de IA en *fact-checking* (1-3)	1.99	.81
Principales limitaciones percibidas en la IA (1=No limita, 2=Limita)	1.47	.50
Confianza en que la IA sustituirá verificadores humanos (1=No, 2=Sí)	1.18	.38

Fuente: Elaboración propia.

La valoración mayoritaria de modelos híbridos y la baja expectativa de sustitución completa de los verificadores humanos sugieren que la aceptación social de la verificación automatizada se articula en torno a criterios de complementariedad más que de automatización plena.

Los resultados recogidos en este bloque (Fase 2) permiten caracterizar las actitudes de la ciudadanía hacia la verificación automatizada como ambivalentes, combinando el reconocimiento de su utilidad operativa con reservas vinculadas a la transparencia, la interpretación contextual y el papel del juicio humano.

5. Conclusiones y Discusión

5.1. Conclusiones

El presente estudio tuvo como objetivo analizar el papel de la inteligencia artificial en los procesos de verificación informativa frente a la desinformación digital, atendiendo de forma integrada a su rendimiento técnico, a la confianza ciudadana y a los retos éticos asociados a su implementación. A partir del análisis de 150 informes de verificación y de una encuesta a 600 usuarios de redes sociales, se obtienen conclusiones que permiten delimitar el alcance real de estas tecnologías en el ecosistema informativo contemporáneo.

En primer lugar, los resultados confirman que la inteligencia artificial se ha consolidado como una herramienta altamente eficaz para la verificación informativa desde el punto de vista operativo. Su capacidad para procesar grandes volúmenes de información con rapidez y para detectar patrones asociados a la desinformación la convierte en un recurso clave frente a la escala y velocidad de los flujos informativos digitales. No obstante, esta eficacia se manifiesta de manera desigual en función del tipo de contenido y del grado de contextualización requerido, lo que evidencia que el rendimiento técnico, por sí solo, no garantiza una verificación plenamente fiable en todos los escenarios.

En segundo lugar, el estudio pone de relieve que la confianza ciudadana en la verificación automatizada se sitúa en niveles moderados y aparece condicionada por factores estructurales como el desconocimiento del funcionamiento de los sistemas, la percepción de opacidad algorítmica y la preocupación por la reproducción de sesgos. Este hallazgo subraya que la aceptación social de la inteligencia artificial en el ámbito informativo no depende únicamente de su precisión, sino de la forma en que sus procesos son comunicados, comprendidos y percibidos como legítimos.

En tercer lugar, y de manera central, los resultados evidencian que los retos éticos constituyen un eje decisivo en la integración de la inteligencia artificial en la verificación informativa. La necesidad recurrente de supervisión humana, las limitaciones en la explicabilidad de las decisiones algorítmicas y las dificultades para auditar los procesos de verificación apuntan a una cuestión ética de fondo: la distribución de responsabilidades en sistemas que influyen directamente en la construcción de la verdad pública. Desde esta perspectiva, la ética no aparece como un complemento normativo, sino como una condición estructural para la legitimidad del *fact-checking* automatizado.

La lectura conjunta de ambas fases permite concluir que los modelos híbridos, que combinan inteligencia artificial y supervisión humana, representan la opción más sólida desde un punto de vista técnico, social y ético. Estos modelos permiten aprovechar la eficiencia de la automatización, al tiempo que incorporan mecanismos de control, contextualización y rendición de cuentas, especialmente relevantes en contenidos complejos o multimodales.

El estudio concluye que la inteligencia artificial constituye un recurso estratégico para reforzar la verificación informativa frente a la desinformación digital, pero solo resulta socialmente aceptable y éticamente sostenible cuando se articula sobre principios de transparencia, explicabilidad, supervisión humana y responsabilidad compartida. Los hallazgos del estudio permiten afirmar que el principal desafío de la verificación automatizada no reside en la mejora incremental de la precisión algorítmica, sino en la construcción de modelos de verificación que sean socialmente comprensibles, éticamente responsables y culturalmente legítimos. El futuro del *fact-checking* automatizado no reside en la sustitución del juicio humano, sino en la construcción de ecosistemas colaborativos en los que la tecnología amplifique —y no sustituya— la capacidad crítica y ética de las personas.

5.2. Discusión

La incorporación de la inteligencia artificial a los procesos de verificación informativa ha introducido transformaciones relevantes en la manera en que se evalúa la fiabilidad de los contenidos en entornos digitales. Más allá de su aportación operativa frente a la desinformación, estos sistemas inciden en los criterios de credibilidad y en las dinámicas de construcción de la confianza informativa, especialmente en un ecosistema caracterizado por la circulación acelerada de información y la creciente presencia de contenidos multimodales.

Los resultados de este estudio confirman que la inteligencia artificial se ha consolidado como un componente estructural de los procesos contemporáneos de verificación informativa, especialmente por su capacidad para operar a gran escala y en tiempos compatibles con la lógica de circulación acelerada de contenidos en redes sociales. El elevado rendimiento técnico observado en el análisis de los informes de verificación coincide con investigaciones recientes que destacan la eficacia de los sistemas automatizados para identificar patrones discursivos, contrastar información y priorizar contenidos potencialmente desinformativos (Guo et al., 2022; Zhou et al., 2022).

No obstante, los hallazgos permiten matizar estas aportaciones al evidenciar que dicha eficacia es condicional y no homogénea. En línea con estudios centrados en verificación automatizada multimodal, los resultados muestran que los sistemas basados en IA presentan mayores dificultades cuando los contenidos requieren interpretación contextual, análisis semántico profundo o detección de manipulaciones visuales complejas (Akhtar et al., 2023; Usman et al., 2023). Esta limitación refuerza la

idea de que el rendimiento técnico no puede evaluarse de forma aislada, sino en relación con el tipo de contenido y el contexto comunicativo en el que se aplica.

La comparación entre los resultados de la Fase 1 y la Fase 2 pone de manifiesto una brecha significativa entre capacidad técnica y confianza ciudadana, fenómeno identificado de forma recurrente en estudios recientes sobre la interacción social con sistemas algorítmicos. A pesar del buen desempeño operativo de la IA, la encuesta revela niveles de confianza moderados, condicionados por el desconocimiento del funcionamiento de los sistemas y por la percepción de opacidad en los procesos de verificación. Esta tensión ha sido ampliamente documentada en la literatura sobre inteligencia artificial responsable, que subraya que la precisión técnica no garantiza legitimidad social si no va acompañada de transparencia y explicabilidad (Bender et al., 2021; Floridi & Cowls, 2022).

Desde una perspectiva ética, los resultados permiten afirmar que los principales retos del *fact-checking* automatizado no se reducen a la existencia de sesgos algorítmicos, sino que se vinculan a problemas más amplios de responsabilidad, gobernanza y rendición de cuentas. La necesidad recurrente de validación humana identificada en los informes analizados apunta a una cuestión ética estructural: la dificultad de atribuir responsabilidades claras cuando las decisiones informativas son mediadas por sistemas automatizados. Este aspecto ha sido subrayado por marcos normativos recientes, que sitúan la supervisión humana y la explicabilidad como principios fundamentales para el uso ético de la inteligencia artificial en contextos sensibles (UNESCO, 2021; Floridi, 2023).

Desde esta perspectiva, la dimensión ética de la verificación automatizada no se limita a la corrección de sesgos o a la transparencia algorítmica, sino que afecta a la definición misma de los criterios mediante los cuales se determina la veracidad informativa en entornos digitales. La integración de sistemas automatizados en el *fact-checking* introduce decisiones implícitas sobre qué contenidos se priorizan, cómo se interpretan y bajo qué parámetros se consideran fiables, lo que sitúa la ética como un componente estructural (y no accesorio) de la verificación informativa basada en inteligencia artificial.

A partir de los resultados obtenidos, este estudio sostiene que la verificación informativa basada en inteligencia artificial no puede interpretarse únicamente como una evolución técnica del *fact-checking*, sino como un proceso que redistribuye la autoridad informativa en el ecosistema digital. La automatización introduce nuevos intermediarios en la validación de la información y desplaza parcialmente los criterios tradicionales de credibilidad, lo que exige repensar la relación entre tecnología, responsabilidad editorial y confianza pública más allá de los marcos puramente operativos.

La preferencia expresada por modelos híbridos en la Fase 2 refuerza estas interpretaciones. Diversos estudios han señalado que los usuarios muestran mayor aceptación de los sistemas automatizados cuando estos operan como apoyo al juicio humano y no como sustitutos autónomos, especialmente en tareas que afectan a la credibilidad informativa y al debate público (Pennycook & Rand, 2022; Tandoc et al., 2023). En este sentido, los resultados de este estudio aportan evidencia empírica que respalda la viabilidad de los enfoques híbridos como solución técnica, social y éticamente más robusta.

Asimismo, los hallazgos dialogan con investigaciones recientes sobre la transformación del *fact-checking* en el ecosistema digital, que advierten de la necesidad de integrar la automatización dentro de marcos de gobernanza informativa más amplios. Estudios centrados en el contexto europeo y latinoamericano subrayan que la legitimidad de la verificación automatizada depende de su articulación con principios de transparencia, responsabilidad editorial y alfabetización mediática (Gutiérrez-Caneda & Vázquez-Herrero, 2024). Los resultados de este trabajo refuerzan esta perspectiva al mostrar que la confianza ciudadana no se construye únicamente sobre la eficacia técnica, sino sobre la comprensión y percepción ética de los sistemas.

Desde una perspectiva educativa y comunicativa, el bajo nivel de conocimiento sobre el funcionamiento de la inteligencia artificial detectado en la muestra confirma que la lucha contra la desinformación no puede abordarse exclusivamente desde soluciones tecnológicas. La literatura reciente ha demostrado que intervenciones centradas en la alfabetización mediática, la reflexión crítica y la atención a la precisión reducen la difusión de contenidos falsos incluso en entornos altamente polarizados (Pennycook & Rand, 2022; Reuters Institute, 2024). En este sentido, los resultados refuerzan la necesidad de integrar la IA en estrategias formativas más amplias, orientadas a fortalecer la competencia mediática de la ciudadanía.

Los resultados analizados ponen de manifiesto que la inteligencia artificial ofrece oportunidades significativas para reforzar la verificación informativa, pero plantea retos éticos y sociales que no

pueden resolverse únicamente mediante mejoras técnicas. La consolidación del *fact-checking* automatizado como práctica legítima depende de su integración en modelos híbridos, transparentes y socialmente comprensibles, capaces de equilibrar eficiencia, responsabilidad y confianza ciudadana.

5.3. Alcance del estudio y proyección de futuras investigaciones

Los resultados de este estudio deben interpretarse atendiendo al marco metodológico en el que se inscriben. En la Fase 1, el análisis se centró en informes de verificación publicados por plataformas que hacen explícito el uso de inteligencia artificial en sus procesos, lo que permitió trabajar con materiales documentados y comparables desde el punto de vista técnico y operativo. Esta elección metodológica responde al objetivo de analizar prácticas verificables y transparentes en el ámbito del *fact-checking* automatizado. Las métricas de rendimiento técnico analizadas deben entenderse como indicadores documentales del desempeño informado por las propias plataformas, adecuados para la caracterización comparativa de enfoques y prácticas de verificación, pero no equivalentes a mediciones experimentales o auditorías externas controladas.

En la Fase 2, la encuesta se diseñó mediante un muestreo por cuotas orientado a captar perfiles diversos de usuarios de redes sociales en contextos hispanohablantes. Esta estrategia, coherente con el carácter descriptivo del estudio, permitió explorar la percepción y la confianza ciudadana en la verificación automatizada en entornos digitales reales, sin pretensión de extrapolación estadística a la totalidad de la población.

A partir de los hallazgos obtenidos, futuras investigaciones podrían ampliar el análisis a otros contextos geográficos y mediáticos, así como incorporar diseños longitudinales que permitan observar la evolución de la confianza y la percepción ética de la inteligencia artificial aplicada al fact-checking. Asimismo, se abren líneas de investigación orientadas a profundizar en el impacto de distintas estrategias de explicabilidad algorítmica, supervisión humana y rendición de cuentas sobre la legitimidad social de estos sistemas.

El estudio apunta a la necesidad de seguir explorando la verificación multimodal y el papel de intervenciones educativas y de diseño de plataformas en el fortalecimiento de la alfabetización mediática y la apropiación crítica de herramientas de verificación automatizada.

Referencias

Akhtar, M., Schlichtkrull, M., Guo, Z., Cocarascu, O., Simperl, E., & Vlachos, A. (2023). Multimodal automated fact-checking: A survey. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 5430–5448. https://doi.org/10.18653/v1/2023.findings-emnlp.361

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Floridi, L. (2023). *Ethics, governance, and policies for artificial intelligence*. Springer. https://doi.org/10.1007/978-3-031-21448-6

Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics, 10*, 178–206. https://doi.org/10.1162/tacl_a_00454

Gutiérrez-Caneda, B., & Vázquez-Herrero, J. (2024). Fact-checking y automatización en el ecosistema informativo digital: retos éticos y profesionales. *Profesional de la Información, 33*(1), e330104. https://doi.org/10.3145/epi.2024.ene.04

Mian, A., & Khan, S. (2020). Coronavirus: The spread of misinformation. *BMC Medicine, 18*, 89. https://doi.org/10.1186/s12916-020-01556-3

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications, 13*, 2333. https://doi.org/10.1038/s41467-022-30073-5

Reuters Institute for the Study of Journalism. (2024). *Digital News Report 2024*. University of Oxford.

Saeidnia, H. R., Hosseini, E., Lund, B., Alipour Tehrani, M., Zaker, S., & Molaei, S. (2025). Artificial intelligence in the battle against disinformation and misinformation: A systematic review of challenges and approaches. *Knowledge and Information Systems, 67*(4). https://doi.org/10.1007/s10115-024-02337-7

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380455

Vallayil, M., Nand, P., Yan, W. Q., & Allende-Cid, H. (2023). Explainability of automated fact verification systems: A comprehensive review. *Applied Sciences, 13*(23), 12608. https://doi.org/10.3390/app132312608

Wardle, C., & Derakhshan, H. (2021). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe.

Zhou, X., Wu, J., & Zafarani, R. (2022). SAFE: Similarity-aware fact checking. *Proceedings of the Web Conference 2022*, 1896–1906. https://doi.org/10.1145/3485447.3512244