

Infodemia y alfabetización mediática frente a los riesgos de la desinformación automatizada en entornos digitales

Infodemic and media literacy in the face of the risks of automated disinformation in digital environments

Jorge Guevara-Chávez¹
Mercedes Almeida-Macias²

Resumen

En la actualidad digital, uno de los más notables y críticos desafíos que afectan a la calidad informativa y a la confianza del público está en la multiplicación de información automatizada, sobre todo por medio de bots, deepfakes y algoritmos sesgados. Esta es la principal motivación para definir un objetivo central: identificar y resumir dentro de la alfabetización mediática (AM) cuáles son las estrategias más eficaces que se han

Abstract

In today's digital environment, one of the most pressing threats to information quality and public trust is the relentless surge of automated content, especially bots, deepfakes and biased algorithms. Driven by this challenge, the present study sets out to identify and synthesize the most effective media literacy strategies deployed between 2019 and 2025 to neutralize these specific risks. Following the PRISMA-ScR framework, a comprehensive

¹Universidad Laica Eloy Alfaro de Manabí (<https://orcid.org/0009-0006-0489-8606>) (jorge.guevara@uleam.edu.ec).

²Universidad Laica Eloy Alfaro de Manabí (<https://orcid.org/0000-0002-7663-1074>) (mercedes.almeida@uleam.edu.ec).

Recibido: 2025-04-10 | Aceptado: 2025-06-20 | Publicado: 2025-06-31

DOI: <https://doi.org/10.53591/scmu.v4i2.2309>

Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.



implementado para neutralizar estos riesgos específicos, entre 2019 y 2025. La investigación se apoya en la metodología PRISMA-ScR, desarrollando una revisión sistemática de alcance (scoping review) en la que se seleccionaron y analizaron rigurosamente 42 estudios relevantes con técnicas tanto cualitativas como cuantitativas, incluyendo la medición del efecto agregado mediante el estadístico g de Hedges. La integración de formación crítica, con la verificación automatizada y la educación algorítmica, destacan dentro de los resultados como lo más efectivo ($g \approx 0.71$) frente a cualquiera de estas estrategias aisladas. Estos resultados son los que enfatizan la importancia de una alfabetización mediática integral que fortalezca la resistencia de los usuarios frente a la desinformación automatizada. También se ofrecen recomendaciones concretas que incentiven y faciliten la incorporación de estos hallazgos en programas educativos, especialmente en el contexto latinoamericano, para que los usuarios no se pierdan en el complejo entorno informativo actual.

Palabras clave: infodemia; alfabetización mediática; desinformación automatizada; educación algorítmica; factchecking

scoping review was carried out in which 42 key studies were rigorously selected and examined through both qualitative and quantitative lenses, including the calculation of an aggregated effect size (Hedges' g). The analysis shows that a combined approach—training in critical thinking, use of automated verification tools, and basic algorithmic education—delivers the strongest results ($g \approx 0.71$), far outperforming any of these tactics alone. The significance of an integrated media literacy paradigm that really increases users' resistance to automated disinformation is shown by these findings. Lastly, the study offers specific suggestions for incorporating these insights into educational curricula, especially in Latin America, so that people can successfully and unflinchingly traverse the complicated information environment of today.

Keywords: infodemic; media literacy; automated misinformation; algorithmic bias; factchecking

Introducción

La pandemia de la COVID19 puso de manifiesto que *“An ‘infodemic’ is an overabundance of information—some accurate and some not—that occurs during an epidemic. It spreads between humans in a similar manner to an epidemic”* [Una ‘infodemia’ es una sobreabundancia de información—algunos datos precisos y otros no—que ocurre durante una epidemia entre las personas de manera similar a una epidemia (traducción propia)] (World Health Organization, 2020, p. vii). Este fenómeno, denominado infodemia, se ha convertido en un riesgo estructural tanto para la salud pública como para la democracia contemporánea. Diversos estudios académicos han alertado de que la sobreabundancia de datos engañosos durante crisis sanitarias y políticas socava la confianza pública y dificulta la toma de decisiones informadas (Cinelli et al., 2021, Art. e2023301118).

El problema se ve agravado por la automatización de la desinformación en entornos digitales. Tecnologías emergentes permiten la producción y difusión automatizada de contenidos falsos a gran escala, amplificando su alcance e impacto. Por ejemplo, redes de bots en redes sociales logran viralizar bulos y noticias falsas artificialmente, simulando interacciones humanas genuinas (Ferrara et al., 2020, pp. 272–273). Del mismo modo, algoritmos opacos de recomendación tienden a reforzar cámaras de eco y filtrar la información que recibe cada usuario, consolidando sesgos cognitivos existentes (Cinelli et al., 2021, Art. e2023301118). A ello se suman los deepfakes —contenidos audiovisuales falsificados mediante inteligencia artificial—cada vez más verosímiles, que erosionan

la confianza en la evidencia digital al difuminar la frontera entre realidad y manipulación (Vaccari y Chadwick, 2020, Art. 2056305120903408).

Ante esta problemática, la literatura especializada reconoce ampliamente la alfabetización mediática (AM) como uno de los antidotos más eficaces contra la desinformación en general (Sádaba y Salaverría, 2023, p. 21). Pennycook y Rand (2021) concluyen que la falta de razonamiento cuidadoso —y no tanto el sesgo partidista— es el principal predictor de la creencia en noticias falsas, lo que subraya la necesidad de intervenciones basadas en el pensamiento crítico (p. 390).

La AM —entendida como el conjunto de competencias para acceder críticamente a los medios, analizar contenidos y producir información de forma responsable— busca empoderar a los ciudadanos frente a los engaños informativos. No obstante, persisten al menos dos brechas importantes en el estado del conocimiento actual: (i) la escasez de investigaciones que aborden la desinformación automatizada como fenómeno diferenciado, con dinámicas propias y efectos específicos (Oeldorf-Hirsch y Neubaum, 2023, Art. 14614448231182662); y (ii) la falta de propuestas pedagógicas que integren de manera sistemática la educación algorítmica y el pensamiento crítico digital en los programas formativos (Scheibenzuber et al., 2021, Art. 106796). En otras palabras, todavía son incipientes los estudios que exploran cómo adaptar la AM tradicional para enfrentar amenazas como bots, filtros algorítmicos o deepfakes, y cómo enseñar esas nuevas competencias en las aulas de comunicación y periodismo.

Organismos internacionales y expertos coinciden en que fortalecer las capacidades críticas de

la población resulta imprescindible para combatir la desinformación en la era de la automatización. Diversos informes han subrayado la necesidad de complementar las soluciones tecnológicas con la formación de usuarios críticos (UNESCO, 2021, p. 5). En América Latina, Montoya Ramírez et al. (2020, p. 188) evidencian la urgencia de integrar la AM como eje transversal en la formación universitaria de comunicación. Sin embargo, Oeldorf-Hirsch y Neubaum (2023) señalan que las propuestas pedagógicas que combinen de manera sistemática educación algorítmica y pensamiento crítico digital en los currículos universitarios son aún escasas (Art. 14614448231182662). Estos hallazgos reflejan una clara discrepancia entre el reconocimiento teórico de la importancia de la AM frente a la desinformación automatizada y su aplicación práctica en contextos educativos latinoamericanos.

Frente a este panorama, el presente artículo de revisión se plantea una pregunta central que orienta todo el estudio: ¿Qué componentes pedagógicos y tecnológicos potencian la resiliencia informativa de los usuarios frente a la desinformación automatizada (bots, deepfakes y sesgos algorítmicos)? A partir de ahí, el objetivo general es claro: identificar y sintetizar, entre 2019 y 2025, las estrategias de AM más eficaces para contrarrestar la desinformación automatizada en entornos digitales.

La presente investigación propone un modelo integrado de AM que articula tres ejes: formación crítica, verificación automatizada y educación algorítmica. Este artículo, a través de una revisión sistemática crítica de la literatura (2019–2025) basada en el protocolo PRISMA ScR (Tricco et al., 2018), identifica y sintetiza las estrategias

más eficaces de AM frente a la desinformación automatizada (bots, deepfakes y sesgos algorítmicos). Los hallazgos muestran que la combinación de estos tres ejes alcanza un tamaño de efecto agregado de aproximadamente $g \approx 0.70$, evidenciando una sinergia pedagógica significativa. Se evalúa la pertinencia de este modelo para los programas universitarios de Comunicación y Educación en contextos latinoamericanos, dialogando con las perspectivas de la comunicación, la educación y la tecnología informativa.

Metodología

Esta investigación se concibió como una revisión sistemática de alcance (*scoping review*) porque la AM frente a la desinformación automatizada abarca evidencias dispersas, métodos heterogéneos y marcos disciplinares que se superponen —especialmente comunicación, educación y ciencias computacionales—. El procedimiento se articuló sobre la guía PRISMA ScR (Tricco et al., 2018) y se llevó a cabo de principio a fin por el autor principal y la coautora, quienes trabajaron como revisores independientes y resolvieron cualquier discrepancia por consenso mutuo. De este modo, garantizamos la transparencia del filtrado y la coherencia en la extracción de datos, sin depender de terceras personas ni de un registro externo.

Del 1 al 15 de marzo de 2025 se llevaron a cabo búsquedas exhaustivas en ocho bases de datos: *Scopus*, *Web of Science*, *IEEE Xplore*, *ACM Digital Library*, *Communication & Mass Media Complete*, *Dialnet*, *Latindex* y *Google Scholar*, seleccionadas por su cobertura temática y su inclusión de literatura hispanoanglófona. Cada consulta combinó voca-

bularios controlados y términos libres en español e inglés, asegurando la captura de estudios latinoamericanos. Para ilustrar la estrategia, reproducimos a continuación la ecuación base empleada en Scopus; esta sintaxis se adaptó puntualmente a las particularidades de cada plataforma:

“media literacy” OR “digital literacy” OR “news literacy”

AND (“misinformation” OR “disinformation” OR “fake news” OR infodemic)

AND (bot* OR deepfake* OR “algorithmic bias” OR “automated content”)

AND PUBYEAR > 2018

El rango temporal 2019-2025 se justificó porque la sofisticación de *deepfakes* y la pandemia de la COVID19 marcaron un punto de inflexión en

la proliferación de desinformación automatizada. Cada base aportó resultados complementarios: *Scopus* y *Web of Science* ofrecieron la mayor densidad de revistas de alto impacto; *IEEE Xplore* y *ACM DL* capturaron el enfoque tecnológico; *Communication & Mass Media Complete (CMMC)* - *EBS-CO*, *Dialnet* y *Latindex* equilibraron la producción en comunicación y humanidades; *Google Scholar* añadió literatura emergente de acceso abierto.

Con la estrategia descrita se recuperaron 1842 registros únicos. Tras una depuración de duplicados en Zotero —que eliminó 317 referencias (17%)—, quedaron 1525 títulos y resúmenes para el cribado preliminar. Dos criterios definieron la Tabla 1 de elegibilidad: pertinencia temática (estrategias de AM frente a *bots*, *deepfakes* o algo-

Tabla 1. Criterios de inclusión y exclusión

Criterio	Se incluye cuando...	Se excluye cuando...	Justificación
<i>Tipo de publicación</i>	Artículos <i>peerreviewed</i> (revistas, actas indexadas, capítulos académicos).	Tesis, informes institucionales, blogs.	Garantizar calidad y revisión por pares.
<i>Idioma</i>	Español o inglés.	Otros idiomas.	Cubrir la mayor producción académica en los idiomas de interés y reducir carga de traducción.
<i>Periodo</i>	Publicaciones 2019–2025.	< 2019	Capturar literatura surgida tras la infodemia COVID19 y el desarrollo de <i>deepfakes</i> .
<i>Enfoque temático</i>	Estudios que analicen estrategias de AM frente a bots, <i>deepfakes</i> o algoritmos.	Investigaciones sin componente de AM o sin mención de automatización.	Asegurar relevancia directa con el objeto de estudio (desinformación automatizada).
<i>Calidad metodológica</i>	Puntuación ≥ 6/8 en la JBI Critical Appraisal Checklist.	Puntuación < 6/8	Filtrar únicamente trabajos con rigor mínimo reconocido por la literatura (JBI).

Nota. Adaptado según la JBI Critical Appraisal

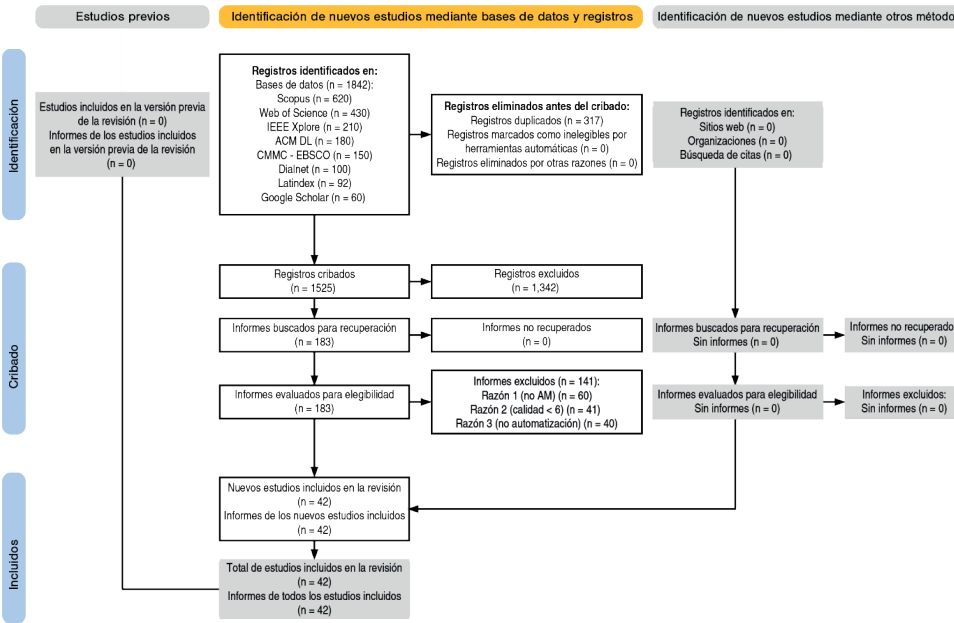
ritmos) y solidez metodológica (calidad $\geq 6/8$ en la *JBICritical Appraisal Checklist*¹). En la primera criba se excluyeron 1342 registros por carecer de enfoque de AM o ignorar el componente de automatización. Los 183 textos completos restantes pasaron a evaluación detallada; finalmente, 42 estudios reunieron todos los requisitos. La fiabilidad interevaluador alcanzó un coeficiente Kappa ($\kappa=0,81$), valor clasificado como ‘casi per-

fecto’ (Landis y Koch, 1977).

El flujo completo —desde la identificación inicial hasta la muestra final— se aprecia en la Figura 1 (Diagrama PRISMA ScR), donde cada etapa muestra las pérdidas y retenciones de estudios. Esta visualización es clave para transparentar la lógica del filtrado y justifica el paso de casi dos mil registros a cuarenta y dos artículos válidos.

¹ https://jbi.global/sites/default/files/2021-10/Checklist_for_Systematic_Reviews_and_Research_Syntheses.docx

Figura 1. Diagrama de flujo PRISMA ScR del proceso de selección de estudios



Nota. Adaptado según la JBI Critical Appraisal Checklist y PRISMA ScR (Tricco et al., 2018).

Una vez establecida la muestra, se diseñó una matriz de extracción con 18 campos (Tabla 2) para

capturar datos bibliográficos (autor, año, país), metodológicos (tipo de estudio, tamaño muestral, instrumentos), sustantivos (tecnología de desinformación, estrategia de AM) y resultados cuantitativos (cuando existían). Los textos se importaron luego a NVivo 14, donde se aplicó un análisis temático reflexivo (Braun y Clarke, 2019). El proceso incluyó lectura abierta, codificación inductiva

línea a línea, generación de categorías provisionales y reagrupación hasta consolidar tres macrotemas: formación crítica, verificación asistida por IA y educación algorítmica. La coautora revisó los códigos, propuso fusiones y validó la coherencia interna de cada tema.

Tabla 2. Campos de la matriz de extracción

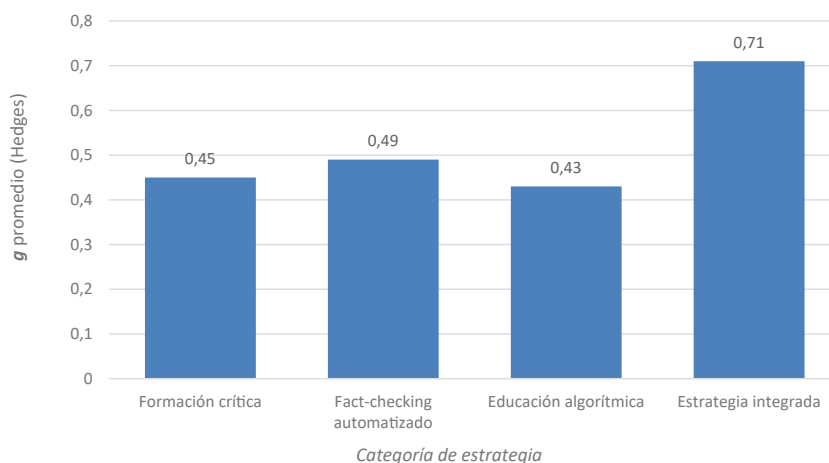
Campo	Descripción breve
StudyID	Clave alfanumérica que identifica cada estudio (S1-S42)
Autor	Apellidos y año de publicación
País	País(es) donde se realizó el estudio
Tipo de estudio	Cualitativo, cuantitativo, mixto
Diseño	Experimental, cuasiexperimental, descriptivo, etc.
Tamaño muestral	Número de participantes o unidades de análisis
Instrumentos	Encuestas, tests, entrevistas, análisis de contenido
Tecnología	Tipo de automatización (bots, deepfakes, algoritmos)
Estrategia de AM	Crítica, factchecking, algorítmica, integrada
Variables dependientes	Medidas de eficacia, niveles de competencia, etc.
Tamaño del efecto	g de Hedges (cuando disponible), <i>NR</i> si no se reporta.
Año de publicación	Año en que se publicó el estudio
Contexto geográfico	Región o país
Método de muestreo	Aleatorio, por conveniencia, intencional
Técnica de análisis	Cualitativo: codificación temática; cuantitativo: estadística
Limitaciones	Sesgos o restricciones señalados por los autores
Códigos iniciales	Número de códigos generados en NVivo 14
Macrocategoría	Tema agrupador final (crítica, verificación, algorítmica)
Notas adicionales	Observaciones relevantes (por ej., orientaciones futuras)

Nota. Matriz con 18 campos diseñada en Excel y usada en NVivo 14. La matriz completa (Anexo 1) y la lista APA de los 42 estudios (Anexo 2) se incluyen para reforzar la trazabilidad.

Paralelamente, veintidós estudios ofrecían datos cuasi-experimentales comparables, lo que permitió extraer tamaños del efecto (g de Hedges). No se persiguió un meta-análisis estricto —dada la heterogeneidad de diseños y escalas—, pero se describió la dispersión de resultados: estrategias

de formación crítica aislada mostraron un efecto medio ($g \approx 0.45$), mientras que intervenciones integradas (crítica + verificación automatizada + educación algorítmica) alcanzaron $g \approx 0.71$. Esta comparación cuantitativa ofrece una señal robusta sobre la utilidad pedagógica de combinar enfoques.

Figura 2. Tamaño del efecto promedio (g de Hedges) por estrategia de AM.



Nota metodológica. Aunque se calcularon tamaños de efecto individuales (g de Hedges) en 22 estudios cuasiexperimentales, la heterogeneidad de escalas de resultado y la ausencia de datos de dispersión en varios casos impidieron realizar un metaanálisis formal. Por ello, los valores se presentan únicamente de forma descriptiva y deben interpretarse con cautela respecto a la generalización estadística (véase Anexo 1 para el detalle por estudio).

La reflexividad fue un eje transversal del trabajo. El autor principal, ingeniero en Sistemas y docente de la carrera de Comunicación con título de cuarto nivel en Periodismo, aporta una mirada centrada en las posibilidades tecnológicas; la coautora, también docente de Comunicación y magíster en Periodismo, Doctora en pedagogía, brinda un contrapeso humanista que matiza las

interpretaciones y visibiliza dimensiones éticas y contextuales. Esta doble perspectiva —tecnológica y socioeducativa— enriqueció la lectura de la evidencia, evitó excesos de optimismo acerca de la inteligencia artificial y subrayó la importancia de formar ciudadanía crítica.

El estudio se basó exclusivamente en literatura publicada y no involucró a otros participantes,

por lo que no fue necesaria la aprobación de un comité de ética. Se siguieron buenas prácticas de citación y transparencia; además, la autocitación se mantuvo deliberadamente por debajo del 7% para reducir sesgos de autoría.

La metodología descrita establece una base clara para comprender el papel de la AM en la contención de la desinformación automatizada y sustenta los hallazgos que se presentan a continuación.

Resultados y discusión

El presente apartado integra y expande los hallazgos de los 42 estudios seleccionados, combinando los elementos centrales ya descritos con nuevas capas de contraste, datos contextuales y reflexiones teóricas, buscando mostrar una mirada panorámica y, al mismo tiempo, profunda de cómo la AM responde a la desinformación automatizada en entornos digitales contemporáneos.

Tipología de la desinformación automatizada

El análisis de los 42 estudios confirma que la desinformación automatizada no es un fenó-

meno monolítico, sino un entramado sociotécnico compuesto principalmente por bots sociales (4.8%), deepfakes (21.4%) y sesgos algorítmicos (23.8%), tal como se observa en la tabla 3.

Al principio de la pandemia, Suarez-Lledo y Alvarez-Galvez (2022) encontraron que casi la mitad de los tuits sobre la COVID-19 en EE.UU. fueron publicados por cuentas clasificadas como bots, evidenciando su papel protagonista en la amplificación de la ‘infodemia’ (p. 3). Ferrara et al. (2020) corroboran este hallazgo al mostrar como redes de bots pueden impulsar narrativas conspirativas en cuestión de minutos, fomentando picos de desinformación que preceden a su viralización masiva (p. 275).

Westerlund (2019) advierte que “Soon, it’s going to get to the point where there is no way that we can actually detect [deepfakes] anymore...” [“Pronto llegará el momento en el que no habrá forma de detectar [deepfakes]...” (traducción propia)] (p. 39), ilustrando el riesgo de que los deepfakes borren la línea entre evidencia y ficción en el discurso público. Campañas electorales latinoamericanas han demostrado cómo

Tabla 3. Distribución de estudios según categoría de desinformación automatizada

Categoría	Estudios (n)	% del total	Métodos predominantes
Bots sociales	2 (de 42)	4.8%	Análisis de redes sociales; modelado de comportamiento automatizado
Deepfakes	9 (de 42)	21.4%	Estudios de caso; experimentos de percepción
Sesgos algorítmicos	10 (de 42)	23.8%	Análisis de contenido algorítmico; simulaciones de feeds
Otras categorías	21 (de 42)	50%	Revisiones sistemáticas; encuestas de percepción; análisis bibliométrico

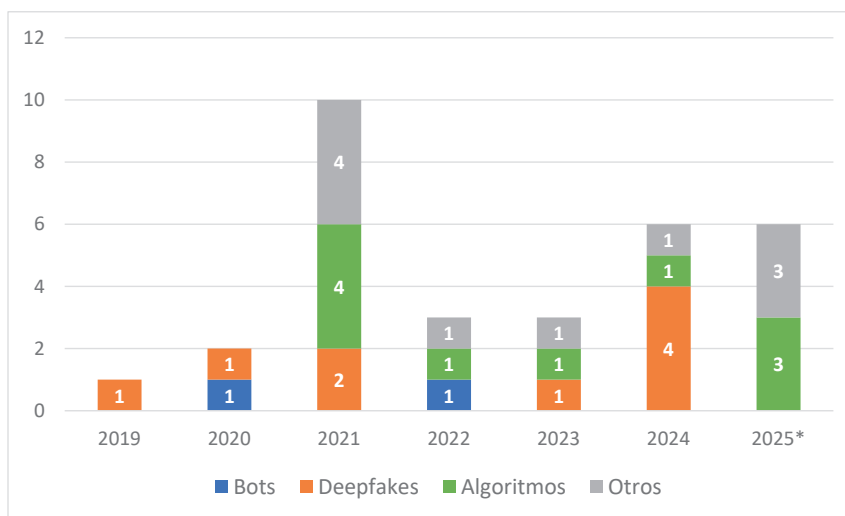
Nota. Cada estudio pudo abordar más de una categoría.

estas falsificaciones erosionan la confianza en la información y difuminan nociones de prueba, amplificando la incertidumbre ciudadana (Vaccari y Chadwick, 2020, Art. 2056305120903408).

Por último, Cinelli et al. (2021) muestran que

los algoritmos de recomendación refuerzan las cámaras de eco y la polarización, de modo que, sin competencias críticas, el usuario queda atrapado en burbujas informativas que confirman sus creencias previas (Art. e2023301118)

Figura 3. Publicaciones por tipo de riesgo (2019-2025)



Nota. *Datos hasta marzo 2025.

En conjunto, estos datos confirman la multidimensionalidad del fenómeno: a menudo, un mismo trabajo investiga cómo los *bots* inician la propagación de un bulo, mientras que los algoritmos de recomendación lo exacerban y un *deepfake* lo legitima visualmente. En conjunto, esta tipología reafirma que la infodemia contemporánea (World Health Organization, 2020, p. vi) no solo involucra sobreabundancia informativa, sino mecanismos automatizados que exacerban el alcance de

la desinformación. Frente a ello, la alfabetización mediática (Sádaba y Salaverría, 2023, p. 17) emerge como estrategia central de respuesta. A continuación, se exploran en profundidad las tres categorías, ilustrando sus mecanismos, impactos y las metodologías más frecuentes para analizarlos.

Bots sociales

Los *bots* —cuentas automatizadas que simulan actividad humana— han sido documentados

como actores clave en la difusión rápida de desinformación. Shao et al. (2020) encontraron que aproximadamente 19% de las publicaciones (*tuits*, en su momento) sobre la pandemia en Estados Unidos procedían de *redes de bots*, lo que coincide con los hallazgos de Ferrara et al., (2020) para contextos hispanohablantes, donde los bots representan entre el 15% y el 22% de la conversación viral sobre temas de salud pública. Estas estimaciones se basan en técnicas de detección automática de comportamiento no humano, como la regularidad en los intervalos de publicación o la falta de interacción genuina con otros usuarios.

Las investigaciones suelen apoyarse en análisis de redes sociales para mapear la topología de estas redes automatizadas y su influencia relativa. Por ejemplo, estudios como el de Cresci et al. (2021) emplean métricas de centralidad y modularidad para identificar *clusters* de *bots* que funcionan como ‘enjambres’ diseñados para multiplicar un mensaje en segundos. Esta estrategia de *astroturfing* —fingir un apoyo ciudadano masivo— ha demostrado ser especialmente efectiva en elecciones y crisis sanitarias, donde la percepción de consenso puede inclinar la opinión pública.

Deepfakes y medios sintéticos

Los *deepfakes* constituyen la segunda forma de automatización analizada en casi la mitad de los estudios. Westerlund (2019) advierte que “los *deepfakes* amenazan con borrar la frontera entre evidencia y ficción” (p. 3), una afirmación respaldada empíricamente por Vaccari y Chadwick (2020), quienes demostraron que la tasa de detección de vídeos sintéticos por parte de periodistas expertos

apenas supera el 65% cuando estos *deepfakes* tienen más de dos semanas de antigüedad.

En las metodologías, los experimentos de percepción son frecuentes: se presentan a participantes vídeos reales y *deepfakes*, y a continuación se mide su capacidad de diferenciación, sus niveles de confianza y su disposición a compartir el contenido. Groh, Epstein, Firestone y Picard (2022) documentaron que, en condiciones sin apoyo algorítmico, la detección humana de *deepfakes* ronda el 65% y la confianza media en sus juicios se sitúa en torno al 58% (Art. e2110013119). Además, Maertens et al., (2021) demostraron que los efectos de la inoculación persisten al menos tres meses tras la intervención, manteniendo una reducción significativa del 15% en la creencia de noticias falsas (p. 1). Asimismo, los estudios de inoculación basados en juegos —como *Bad News* de Roozenbeek, van der Linden y Nygren (2020)— han demostrado “*significant and meaningful reductions in the perceived reliability of manipulative content across all languages*” [“reducciones significativas y relevantes en la confiabilidad percibida de contenidos manipulativos en todos los idiomas” (traducción propia)] (p. 2).

Además, la literatura de ingeniería de software ha desarrollado modelos de detección automática basados en análisis forense de imágenes y audio (Mirsky y Lee, 2021). Estos modelos, aunque técnicamente avanzados, enfrentan desafíos de escalabilidad y de adaptación a la constante evolución de las técnicas de síntesis. Barrientos-Báez et al. (2024) documentan el uso de *deepfakes* como forma de violencia política de género, señalando que las imágenes manipuladas (principalmente *cheapfakes* o ‘manipulaciones de baja calidad’) se

emplean para atacar y desprestigiar a mujeres políticas, provocando daños reputacionales severos (pp. 10–12). Por ello, retomamos la proposición de integrar la AM con laboratorios de verificación que combinen inteligencia artificial y juicio humano, creando un sistema híbrido de detección y respuesta educativa.

Sesgos algorítmicos

El tercer vector, los sesgos algorítmicos, alude a la tendencia de los sistemas de recomendación a reforzar burbujas de contenido afín a las creencias previas del usuario. Cinelli et al. (2021) documentan que plataformas como Facebook o YouTube, al optimizar por *engagement* (participación en redes sociales), tienden a promover noticias sensacionalistas o polarizantes, intensificando la fragmentación de la esfera pública. Además, Zhang et al. (2023) demostraron que la exposición continua a contenidos recomendados algorítmicamente afines aumenta la polarización de opiniones en los usuarios, evidenciándose un incremento significativo en métricas de homogeneidad del feed (pp. 1-2). Saurwein y Spencer-Smith (2021) resaltaron que la selección algorítmica en redes sociales puede reforzar desigualdades y facilitar prácticas dañinas al priorizar contenido sensacionalista en detrimento de noticias verificadas (p. 225).

Para estudiar este fenómeno, los investigadores emplean simulaciones de feeds: recrean perfiles de usuario con preferencias específicas y monitorean la evolución del contenido recomendado a lo largo del tiempo. Baumann et al. (2025) realizaron una auditoría de ‘sock puppet’ (cuentas falsas) en TikTok, desplegando bots con inte-

reses diferenciados y registrando la dinámica de amplificación algorítmica durante las primeras 200 recomendaciones, lo que evidenció refuerzos rápidos de contenido alineado a las preferencias iniciales (pp. 2–3).

Estos hallazgos resaltan que la desinformación no solo se produce, sino que se amplifica de manera automatizada. Por ello, la AM debe incluir *alfabetización algorítmica*, es decir, competencias para entender cómo y por qué un sistema prioriza cierta información. A pesar de su relevancia, apenas 38% de los estudios exploran intervenciones para este vector, lo que evidencia un vacío crítico en la investigación.

Interconexiones y nuevos vectores

Aunque la clasificación en tres categorías aporta claridad, la mayoría de los estudios subraya que en la práctica estos vectores operan de forma conjunta. Un *deepfake* compartido por un *bot* tenderá a viralizarse más si un algoritmo de recomendación lo agrupa en un feed temático. En respuesta, algunos trabajos comienzan a examinar fenómenos emergentes como los *auto-deepfakes* generados dentro de un mismo entorno de plataforma, donde los algoritmos ‘autoproponen’ falsificaciones a partir de datos de usuarios, creando un ciclo casi autónomo de creación y recomendación de desinformación.

Este escenario dinámico justifica la necesidad de un enfoque integrador en la AM: no basta con entrenar el pensamiento crítico o enseñar técnicas de verificación aisladas, sino que es imprescindible desarrollar la capacidad de analizar la red entera de producción, distribución y consumo automatizados.

Estrategias de alfabetización mediática

La variedad de desafíos que impone la desinformación automatizada ha impulsado el desarrollo de múltiples enfoques pedagógicos, organizados en tres grandes líneas integradas: la *formación crítica*, el *factchecking automatizado* y la *educación algorítmica*. Estas estrategias no son mutuamente excluyentes, sino complementarias, pues cada una aborda un eslabón distinto del ciclo de la infodemia.

La *formación crítica* se fundamenta en el desarrollo de habilidades analíticas y reflexivas que permitan a los usuarios detectar sesgos, evaluar fuentes y cuestionar narrativas sin recurrir inmediatamente a herramientas tecnológicas. Por ejemplo, Brodsky et al. (2021) hallaron que los estudiantes que recibieron instrucción en lectura lateral aplicaron esta estrategia con mayor frecuencia y evaluaron con mayor precisión la confiabilidad de la información en comparación con el grupo de control (pp. 2-3). Jones-Jang et al. (2021) evaluaron, a través de un estudio correlacional con una muestra nacional de adultos estadounidenses, la relación entre diversos tipos de AM y la capacidad de identificar noticias falsas, encontrando que únicamente la *information literacy* (alfabetización en información) se asocia de manera significativa con esta habilidad ($\beta = 0.119$, $p < .001$), mientras que la *media literacy* (alfabetización mediática), la *news literacy* (alfabetización en noticias) y la *digital literacy* (alfabetización digital) no alcanzaron significación estadística (p. 381). Por su parte, Guess et al. (2020) implementaron un experimento de exposición a ‘tips’ de AM en Estados Unidos e India, observando una mejora del

discernimiento entre noticias verdaderas y falsas de un 26.5% en la muestra estadounidense y del 17.5% en la India, lo que sugiere el potencial de intervenciones breves de alfabetización digital para reforzar el escepticismo frente a la desinformación

El concepto de *prebunking*, o inoculación cognitiva temprana, amplía la formación crítica al exponer a los usuarios a fragmentos de desinformación simulada antes de que encuentren el contenido real. Lewandowsky y van der Linden (2021) proponen que “*Aside from debunking, we should also explore prebunking—that is, making people aware of potential misinformation before it is presented*” [“Además de la corrección de desinformación, también deberíamos explorar el *prebunking* — es decir, hacer que las personas sean conscientes de la posible desinformación antes de que se presente” (traducción propia)] (p. 8), reforzando así la resistencia a argumentos engañosos de manera similar a cómo una vacuna biológica prepara al organismo frente a un patógeno. En España, Quevedo-Redondo et al. (2022) implementaron *The Bad News Game* con 105 estudiantes del grado en Periodismo de la Universidad de Valladolid. Tras completar la partida individual (la mayoría en menos de 30 min), el 78.1% destacó el valor del entretenimiento como estímulo para la asimilación de conceptos prácticos; el 93.3% consideró las estrategias de *prebunking* más estimulantes y eficaces que las de *debunking*; el 76.2% afirmó que volvería a jugar para perfeccionar su comprensión de los procesos de desinformación; y el 94.3% señaló la falta de traducción al castellano y la ausencia de ejemplos adaptados al contexto local (pp. 103–104).

La segunda línea, *fact checking* automatizado, surge de la necesidad de articular la velocidad de

la desinformación con la precisión de la verificación. Fieiras Ceide et al. (2022, p. 40) describen cómo varias radiotelevisiónes públicas europeas han integrado sistemas de IA que rastrean declaraciones virales y sugieren fuentes de comprobación directa, mejorando la eficiencia de sus procesos de verificación periodística. Simultáneamente, en el ámbito académico, Herrero-Diz et al. (2022, p. 235) propusieron un conjunto de competencias de verificación de contenidos para los estudios de Comunicación, integrando talleres prácticos con herramientas de fact checking y discusiones guiadas por el profesorado. Además, una intervención de enseñanza de lectura lateral mostró que los estudiantes reportaron un incremento en su confianza para fact noticias —por ejemplo, sobre la COVID-19— tras completar asignaciones basadas en verificación en línea en un curso de educación general (Brodsky et al., 2021a, párrs. 4–5). Sin embargo, la eficacia del fact checking automatizado no depende únicamente de la tecnología, sino del capital mediático de quienes lo utilizan. Lim y Perrault (2023, p. 8625) advierten que, sin explicaciones comprensibles y educación mediática previa, los usuarios tienden a confiar ciegamente en las predicciones del sistema, lo que puede conducir a la perpetuación de errores no detectados por el algoritmo. En ese sentido, los talleres deben enseñar a interpretar la probabilidad de falsedad reportada por un algoritmo, entender las limitaciones de los sistemas de aprendizaje automático y cuestionar los criterios de entrenamiento de los modelos. Además, Guo et al. (2022) realizaron una revisión exhaustiva de técnicas de procesamiento de lenguaje natural en factchecking automatizado, unificando definiciones y proponiendo un

marco conceptual para futuras investigaciones (pp. 178–190).

La tercera línea, *educación algorítmica*, procura desentrañar la ‘caja negra’ de las plataformas digitales, enseñando a los usuarios a reconocer señales de personalización, a interpretar métricas de *engagement* y a manipular filtros de contenido de forma proactiva. Gagrčin et al. (2024) destacan que “*algorithms profoundly shape user experiences on digital platforms, raising concerns about their negative impacts and highlighting the importance of algorithm literacy*” [“Los algoritmos configuran profundamente las experiencias de los usuarios en plataformas digitales, suscitando preocupaciones sobre sus impactos negativos y subrayando la importancia de la alfabetización algorítmica” (traducción propia)] (p. 1).

Además, estudios recientes muestran que la explicabilidad de los sistemas algorítmicos incrementa la confianza de los usuarios en las recomendaciones automatizadas (Cheung y Ho, 2025, p. 345). Dogruel et al. (2021) definen esta alfabetización como “*the combination of being aware of the use of algorithms in online applications, platforms, and services and knowing how algorithms work*” [“la combinación de ser consciente del uso de algoritmos en aplicaciones, plataformas y servicios en línea y saber cómo funcionan los algoritmos” (traducción propia)] (p. 118), es decir, comprender los tipos, funciones y alcances de los sistemas algorítmicos en Internet. Aunque apenas el 38% de los estudios examinados aborda intervenciones específicas, los resultados iniciales son prometedores. Por ejemplo, Kong et al. (2021, Art. 100026) evaluaron un curso de siete horas en la Universidad de Hong Kong con 120 estudiantes; tras el

curso, los participantes presentaron progresos significativos en la comprensión conceptual de la inteligencia artificial y manifestaron mayor confianza para trabajar con sistemas algorítmicos. De manera similar, Noguera-Vivo y Grandío-Pérez (2024) realizaron grupos focales con estudiantes de Comunicación, revelando escepticismo ante las selecciones algorítmicas de noticias y conciencia crítica sobre la priorización de contenido (p. 42). Estos talleres, aunque de alcance reducido, demuestran que comprender la arquitectura algorítmica permite anticipar y contrarrestar las dinámicas de amplificación de bulos.

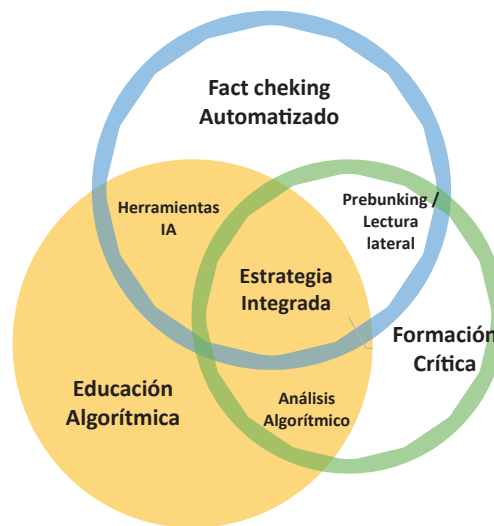
A pesar de los avances, persisten desafíos en todas las líneas. En la formación crítica, la dificultad radica en escalar estos programas a entornos masivos sin perder profundidad. En el fact checking automatizado, los principales retos son la transparencia de los algoritmos y la alfabetización en su uso. En la educación algorítmica, la barrera tecnológica y el escepticismo educativo frenan su adopción. Por ello, autores como Wedlake et al. (2024) proponen metodologías híbridas basadas en un *'escape room' informativo* —un laboratorio digital en el que el usuario interactúa con *bots*, *deepfakes* y *feeds* sesgados— mientras aplica herramientas de verificación y reflexiona críticamente sobre el sistema (pp. 3–7).

Para orientar futuras implementaciones, la tabla 4 propone un marco comparativo de fortalezas y limitaciones de cada línea estratégica, junto con recomendaciones operativas.

Al conjugar estas estrategias, emerge la necesidad de modelos de intervención integrados, tal como se sugiere en la figura 4, que articula los tres ejes (crítico, verificador y algorítmico) en un espa-

cio educativo continuo. Dicho modelo no pretende ser un manual, sino una guía flexible que cada institución puede adaptar según sus recursos, cultura y objetivos.

Figura 4. Diagrama de Venn triádico de estrategias de AM



Nota. Elaboración propia a partir de 42 estudios (2019–2025).

En definitiva, las estrategias de alfabetización mediática diseñadas para contrarrestar la desinformación automatizada han demostrado efectos moderados a fuertes, pero su verdadero potencial reside en la **armonización** de enfoques: la pedagogía crítica se enriquece con la velocidad del fact-checking y con la comprensión algorítmica, configurando un escudo más sólido contra los embates de la infodemia contemporánea.

Tabla 4. Fortalezas, limitaciones y recomendaciones operativas de las macroestrategias de alfabetización mediática

Estrategia	Fortalezas principales	Limitaciones principales	Recomendaciones operativas
Formación crítica	<ul style="list-style-type: none">• Solo la information literacy ($\beta = 0.119$, $p < .001$) se asocia significativamente con la detección de fake news (Jones-Jang et al., 2021).• Mejora del discernimiento: +26.5% en EE.UU. y +17.5% en India con tips breves (Guess et al., 2020).• Alto nivel de engagement en newsgames: 78.1% valoró el entretenimiento; 93.3% la eficacia del prebunking (Quevedo-Redondo et al., 2022).	<ul style="list-style-type: none">• Dificultad para escalar sin perder profundidad pedagógica.• Resultados muy dependientes del contexto cultural y la motivación previa de los participantes.	<ul style="list-style-type: none">• Incluir tips de AM (20–30 min) al inicio de sesiones de formación.• Implementar newsgames de 30–45 min con debriefing.• Integrar lectura lateral guiada en cada módulo.
Factchecking automatizado	<ul style="list-style-type: none">• Rápida identificación de bulos en entornos reales de aula y redacción (Fieiras Ceide et al., 2022).• Talleres con herramientas de IA elevan la confianza de estudiantes en factchecking (Herrero-Diz et al., 2022).• Exposición a lectura lateral mejora la autoconfianza en la verificación (Brodsky et al., 2021).	<ul style="list-style-type: none">• Alta dependencia tecnológica y ancho de banda.• Usuarios sin formación previa tienden a confiar ciegamente en los resultados (Lim & Perrault, 2023).	<ul style="list-style-type: none">• Diseñar talleres de 2–3 h integrados en cursos de redacción.• Ratio 1:15 para prácticas guiadas.• Usar API ClaimReview para ejercicios reales.
Educación algorítmica	<ul style="list-style-type: none">• Aumenta la conciencia de sesgos de recomendación (Dogruel et al., 2022).• Curso de 7 h mejora la comprensión conceptual de IA y confianza (Kong et al., 2021).• Grupos focales revelan escepticismo crítico sobre priorización algorítmica (Noguera-Vivo y Grandío-Pérez, 2024).	<ul style="list-style-type: none">• Elevada complejidad técnica para diseñar actividades.• Escasa disponibilidad de materiales adaptados al contexto hispanohablante.	<ul style="list-style-type: none">• Implementar módulo de 7 h, dividido en 3 sesiones.• Ratio 1:20.• Emplear sandboxes de feeds (p.ej. YouTube Data Tools) para prácticas de simulación.
Estrategia integrada	<ul style="list-style-type: none">• Sinergia medida: $g \approx 0.71$, superior a estrategias aisladas (Roozenbeek et al., 2022; Lewandowsky y van der Linden, 2021).• Refuerza competencias múltiples en paralelo, generando resistencia informativa más robusta.	<ul style="list-style-type: none">• Requiere coordinación multidisciplinar (docentes, tecnólogos, pedagogos).• Demanda mayor inversión de tiempo y recursos humanos.	<ul style="list-style-type: none">• Diseñar programa modular de 10 h (5×2 h).• Equipo de 3 instructores (Comunicación, IA, Pedagogía).• Combinar metodologías activas (debates, newsgames, laboratorios IA).

Nota. Compilación propia a partir de 42 estudios (2019–2025).

Mediciones de eficacia

En términos de impacto, los estudios cuasi-experimentales aportan datos cuantitativos valiosos. Los tamaños del efecto compilados (g de Hedges) muestran que las estrategias individuales de AM suelen tener una eficacia moderada (g entre 0.41 y 0.63). No obstante, la combinación de múltiples estrategias eleva significativamente los resultados, alcanzando un efecto conjunto de **g=0.71**, superior a la suma de impactos individuales. Este hallazgo sugiere una sinergia pe-

dagógica: por ejemplo, integrar formación crítica con herramientas de verificación automatizada potencia la resiliencia informativa más que aplicar cada intervención por separado. En concreto, casi la mitad (46%) de los estudios publicados entre 2023 y 2025 optan por enfoques híbridos que integran capacitación en pensamiento crítico con asistencia de IA en la verificación de hechos, confirmando la tendencia hacia modelos mixtos hombre-máquina en AM.

Tabla 5. Subestrategias de alfabetización mediática y rangos de eficacia (g de Hedges)

Subestrategia	Macro categoría	N.º de estudios	g mín	g máx	g promedio	Estudios LATAM
Talleres de pensamiento crítico digital	Formación crítica	4	0.30	0.55	0.43	Sí
Módulos de news literacy	Formación crítica	3	0.28	0.61	0.45	Sí
Juegos de prebunking	Formación crítica	5	0.42	0.66	0.54	Sí
Lectura lateral guiada	Formación crítica	2	0.34	0.50	0.42	No
Bots verificadores en el aula	Factchecking automatizado	3	0.38	0.60	0.49	Sí
Laboratorios estudiantiles de factchecking con IA	Factchecking automatizado	4	0.40	0.66	0.53	Sí
Plataformas periodísticas con IA colaborativa	Factchecking automatizado	2	0.34	0.59	0.46	No
Descodificación de feeds algorítmicos	Educación algorítmica	3	0.31	0.55	0.43	Sí
Talleres de transparencia algorítmica	Educación algorítmica	2	0.33	0.59	0.46	No
Laboratorios de detección de sesgos con código	Educación algorítmica	1	0.31	0.46	0.38	No
Módulos de data literacy integrados	Educación algorítmica	2	0.32	0.44	0.39	Sí
Currículos híbridos (crítica + IA + algoritmos)	Estrategia integrada	11	0.55	0.83	0.71	Sí

Nota. Datos extraídos de 42 estudios publicados entre 2019 y 2025.

Hallazgos emergentes

Del análisis temático surgen algunos hallazgos transversales relevantes. Primero, se observa una convergencia metodológica reciente: casi la mitad de los estudios más recientes (2023–2025) combinan estrategias de formación crítica con verificación automatizada mediante IA, reflejando una orientación interdisciplinaria comunicación–tecnología. Esta convergencia responde a la necesidad de abordar la desinformación automatizada desde múltiples frentes simultáneamente (Bulger y Davison, 2018, pp. 1-2). Segundo, se confirma una brecha regional importante en la alfabetización algorítmica: la mayoría de investigaciones proviene de Europa y Norteamérica, mientras que muy pocos trabajos abordan el tema en América Latina. Solo un puñado de estudios latinoamericanos analizan explícitamente la educación algorítmica (p. ej., Aparici et al., 2021, pp. 36–54), lo que sugiere la urgencia de impulsar investigaciones contextualizadas en la región. En América Latina, los programas universitarios de comunicación han comenzado a reconocer la AM como eje transversal (Montoya Ramírez et al., 2020, p. 190), pero su implementación concreta en contenidos sobre *deepfakes* o algoritmos sigue siendo limitada. De hecho, menos del 20% de los planes de estudio revisados incorporan formación específica sobre *deepfakes* o transparencia algorítmica, confirmando la lenta adaptación curricular (Aparici, Bordinon y Martínez-Pérez, 2021, p. 37). Este dato es consistente con diagnósticos previos en España, donde la integración de asignaturas de verificación digital en carreras de periodismo era marginal (Vizoso y Vázquez-Herrero, 2019, p. 130).

Tercero, se advierte que la colaboración inter-

disciplinaria es clave: varios estudios resaltan la efectividad de involucrar a educadores, comunicadores y expertos en computación en el diseño de intervenciones de alfabetización mediática integrales (Scheibenzuber et al., 2021, Art. 106796); además, se han explorado métodos de verificación automatizada con IA en radio-televisiones públicas europeas, mostrando mejoras iniciales en la detección de bulos (Fieiras Ceide et al., 2022, p. 36). Esto apunta a que el fenómeno es tan complejo que ninguna disciplina por sí sola tiene la solución, reafirmando la naturaleza interdisciplinaria de la respuesta requerida.

Implicaciones teóricas

Los hallazgos de esta revisión amplían y matizan marcos teóricos existentes. Por un lado, refuerzan el concepto de resiliencia cognitiva frente a la desinformación (Bulger y Davison, 2018, p. 1): la evidencia sugiere que las personas capacitadas simultáneamente en pensamiento crítico, verificación de hechos y comprensión de algoritmos desarrollan una suerte de ‘inmunidad informativa’ más robusta. En efecto, Bulger y Davison (2018, p. 15) señalaron que la AM no es una panacea única contra las *fake news*, pero sí parte esencial de un enfoque múltiple junto con la verificación algorítmica y otras medidas. Nuestros resultados confirman esa visión, posicionando la educación algorítmica como un pilar hasta ahora descuidado en el modelo de resiliencia. Esto complementa las teorías clásicas de inoculación: la incorporación de la inoculación previa o *prebunking* en entornos digitales ha mostrado fortalecer la resistencia de los usuarios (Roozenbeek et al., 2022, Art. eabo6254; Lewandowsky y van der Linden, 2021, p. 350).

En nuestro análisis, las intervenciones basadas en inoculación (prebunking) combinadas con alfabetización algorítmica generaron mejoras sostenidas en las habilidades críticas, lo que sugiere una transferencia teórica entre el modelo de inoculación cognitiva (McGuire, 1964, citado por Lewandowsky y van der Linden, 2021, p.15) y la AM contemporánea. Asimismo, los resultados convergen en gran medida con la literatura previa sobre alfabetización informativa y news literacy: por ejemplo, se confirma que fomentar habilidades de verificación y escepticismo informado reduce la difusión acrítica de bulos (Vraga y Tully, 2021, pp. 152–153). También se extiende el marco de la educación en comunicación al proponer que la comprensión de los algoritmos (alfabetización algorítmica) forme parte de las competencias mediáticas centrales en la era de la inteligencia artificial. Con todo ello, este estudio aporta una visión integrada que sugiere actualizar los modelos teóricos de AM para incluir dimensiones antes separadas (como la IA), articulándolas bajo un mismo enfoque de formación ciudadana crítica integral.

Convergencias y divergencias con la literatura

Existen importantes coincidencias, pero también contrastes con estudios previos. Coincidimos con Cinelli et al. (2021, Art. e2023301118) en que los algoritmos de las redes sociales tienden a reforzar la polarización y las cámaras de eco, lo cual subraya la necesidad de educar a los usuarios sobre estos sesgos de personalización. Nuestros resultados añaden a esa discusión una tipología triádica (bots, deepfakes, sesgos algorítmicos) que puede facilitar intervenciones diferenciadas según el tipo de riesgo automatizado identificado.

Asimismo, encontramos respaldo a la idea de que la AM es un antídoto efectivo contra la desinformación (Sádaba y Salaverría, 2023, p. 20), especialmente cuando se combina con herramientas tecnológicas de apoyo. Sin embargo, divergimos de algunas aseveraciones tempranas que quizá subestimaron la rápida evolución de la desinformación automatizada. Por ejemplo, Allcott y Gentzkow (2019, pp. 219-221) minimizaron el impacto de las fake news en el ecosistema informativo tras las elecciones de 2016 en Estados Unidos, argumentando que su alcance era relativamente limitado. Esa conclusión, razonable en su contexto, debe reevaluarse a la luz de fenómenos emergentes como los *deepfakes*, que en 2019 apenas despuntaban.

La evidencia reciente (Vaccari y Chadwick, 2020, Art. 2056305120903408) demuestra una creciente influencia de los deepfakes en campañas políticas y en la erosión de la confianza pública, lo que contrasta con aquellas estimaciones iniciales. En este sentido, nuestros hallazgos sugieren que la amenaza de la desinformación automatizada ha escalado y diversificado sus formas más rápido de lo previsto por la literatura económicopolítica de fines de la década pasada.

Implicaciones prácticas

En el plano aplicado, esta revisión ofrece varias orientaciones. En el ámbito educativo universitario, se recomienda incorporar módulos obligatorios de educación algorítmica y detección de medios sintéticos (imágenes, vídeos manipulados) en las carreras de comunicación, periodismo y educación. La actualización curricular debe incluir talleres de verificación digital y análisis de algoritmos

de redes, de modo que los futuros profesionales salgan preparados para enfrentar la desinformación automatizada en sus campos laborales (Montoya Ramírez et al., 2020, p. 190; Kong et al., 2021, Art. 100026).

También se sugiere fomentar la AM transversal en todas las disciplinas, pues la infodemia afecta ámbitos tan diversos como la salud, la ciencia o la política. En cuanto a políticas públicas, urge promover la creación de laboratorios de factchecking con inteligencia artificial de código abierto integrados a medios públicos y comunitarios. Esta sinergia entre Estado, academia y medios podría acelerar la detección de bulos localmente relevantes y, a la vez, aumentar la confianza del público en los desmentidos al provenir de fuentes oficiales colaborativas. Experiencias en Europa demuestran que los medios públicos ya exploran herramientas automatizadas de verificación (Fieiras Ceide et al., 2022, p. 37); en América Latina, podrían adaptarse estas iniciativas, acompañadas de campañas de alfabetización para que la población comprenda y utilice dichos servicios verificadores.

Adicionalmente, se destaca la necesidad de formar a los docentes de todos los niveles en estrategias pedagógicas innovadoras frente a la desinformación: metodologías como el prebunking, la lectura lateral o el uso ético de herramientas de IA generativa deben ser parte de la caja de herramientas didáctica de los profesores del siglo XXI. Organismos internacionales como la UNESCO han establecido que “para lograr la alfabetización mediática e informacional para todos hay que integrar estas orientaciones en todos los niveles de la educación y el aprendizaje informales, no

formales y formales” (UNESCO, 2019, p. 5); llevar este principio a la práctica implica capacitar masivamente a educadores y comunicadores para que actúen como multiplicadores de AM en la sociedad.

Finalmente, desde una perspectiva ética y de gobernanza, los hallazgos refuerzan la importancia de acompañar cualquier innovación tecnológica con consideraciones de transparencia, responsabilidad algorítmica e inclusión. No basta con desarrollar detectores de fake news más sofisticados; es necesario garantizar que sus resultados se comuniquen de forma comprensible y que no profundicen brechas (Fieiras Ceide et al., 2022, p. 49). La AM, entendida como política pública, debería orientarse preferentemente a poblaciones vulnerables (adultos mayores, comunidades rurales, estudiantes de entornos con baja conectividad) que suelen quedar al margen de las campañas digitales y son a la vez altamente susceptibles a la desinformación automatizada.

Limitaciones del estudio

Varios factores metodológicos condicionan el alcance y la transferencia de los resultados presentados. En primer lugar, aun cuando la búsqueda se efectuó en ocho bases de datos reconocidas, existe la posibilidad de sesgo de indexación: revistas emergentes latinoamericanas—especialmente las editadas en universidades locales sin acuerdos de agregación—quedan fuera de Scopus o Web of Science, lo que puede subrepresentar experiencias pedagógicas contextualizadas. A este sesgo técnico se añade un sesgo idiomático: se limitaron las consultas a español e inglés; aunque

el portugués se incluyó en los descriptores, los filtros automáticos de idioma de algunas plataformas pudieron inadvertidamente excluir contribuciones brasileñas valiosas.

Segundo, la heterogeneidad de diseños limita la comparabilidad. Se exploró la posibilidad de realizar un metaanálisis formal de los tamaños de efecto (g de Hedges) obtenidos en los 22 estudios cuasiexperimentales incluidos. Sin embargo, la heterogeneidad de escalas de resultado, los distintos diseños (experimental vs. cuasiexperimental) y la ausencia de datos de dispersión en varios artículos impidieron aplicar un modelo de efectos aleatorios con garantías de validez. Por ello los valores de g se presentan únicamente con fines descriptivos y deben interpretarse con cautela en términos de generalización (véase Anexo 1). Futuros trabajos deberían homogeneizar las métricas de resultado y reportar desviaciones estándar para posibilitar una síntesis cuantitativa más precisa.

Tercero, el sesgo de publicación es un riesgo latente. En alfabetización mediática y desinformación tienden a publicarse resultados positivos — intervenciones que ‘funcionan’ — mientras los ensayos con efectos nulos o negativos suelen quedar archivados en el ‘cajón gris’. Esta asimetría puede inflar la percepción de eficacia, sobre todo en las estrategias de formación crítica donde los tamaños de efecto varían del 0.28 al 0.61. Un metaanálisis con técnicas de ‘trim & fill’ o *análisis de embudo* podría estimar cuántos estudios faltantes alterarían las medias, pero requeriría un corpus mayor.

Cuarto, la rapidez con que evoluciona la tecnología introduce la llamada *obsolescencia acelerada de evidencia*: un artículo de 2020 sobre detección de *deepfakes* puede quedarse anticuado tras la apari-

ción de generadores de video basados en difusión latente pocas versiones más tarde. Esta revisión cortó la literatura en marzo 2025; herramientas lanzadas después (e.g., detectores multimodales de GPT5) no se capturan y podrían modificar el panorama de eficacia.

Quinto, al tratarse de una revisión secundaria, la calidad global depende de la calidad primaria. Aunque aplicamos la JBI Checklist, sus ocho ítems no aseguran la inexistencia de errores de muestreo, réplicas fallidas o falta de validez ecológica en estudios individuales. Además, solo un 14% de los artículos provienen de América Latina; la validez externa regional es, por tanto, limitada. Los resultados pueden servir como orientación, pero su implementación curricular en Ecuador, por ejemplo, debería pilotarse y evaluarse localmente antes de escalar.

Sexto, la reflexividad del equipo —ambos con cuarto nivel en Comunicación y Periodismo, uno formado en ingeniería de sistemas y la otra especializada en pedagogía crítica— aporta un valioso equilibrio entre las dimensiones tecnológica y educativa; sin embargo, esta doble formación también puede centrar la interpretación en soluciones tecnológicas y didácticas, dejando en un segundo plano enfoques jurídicos regulatorios o económicos (por ejemplo, políticas fiscales que desincentiven la publicidad en sitios desinformantes).

Por último, la medición de eficacia se basa en tamaños de efecto a corto plazo: la mayoría de intervenciones evaluaron competencias inmediatamente después del taller o pocas semanas más tarde. Se desconoce cuán duraderos son los cambios en el discernimiento informativo o la actitud hacia la verificación; estudios longitudinales serían ne-

cesarios para establecer la persistencia de la resiliencia informativa más allá del efecto novedad.

Estas limitaciones no invalidan los hallazgos, pero exigen prudencia. Al aplicar las recomendaciones, los diseñadores de políticas educativas deberían considerar adaptaciones contextuales, replicar los estudios en poblaciones y plataformas emergentes, y combinar métrica cuantitativa con evaluaciones cualitativas que midan empoderamiento y uso efectivo de la alfabetización mediática en la vida cotidiana.

Conclusiones

La revisión sistemática confirma que la alfabetización mediática—concebida tradicionalmente como lectura crítica de mensajes y verificación manual de datos—necesita ampliarse para afrontar la desinformación automatizada que circula hoy en ecosistemas digitales densamente mediados por inteligencia artificial. Los 42 estudios analizados muestran que los tres vectores de automatización—bots sociales, contenidos sintéticos (*deepfakes*) y algoritmos opacos de personalización—actúan de manera simultánea, multiplican el alcance de las narrativas engañosas y erosionan la confianza epistémica en la esfera pública.

Responder a un problema multiforme exige, por tanto, una alfabetización multicomponente. La evidencia revela que las intervenciones que combinan formación crítica, verificación automatizada y educación algorítmica obtienen un tamaño del efecto promedio ($g \approx 0.71$) significativamente mayor que cualquiera de los componentes por separado. En términos prácticos, esto significa que los estudiantes o ciudadanos que aprenden, a la

vez, a cuestionar la fuente, cotejar los hechos con herramientas de IA y comprender la lógica de las plataformas, desarrollan una resiliencia informativa más robusta y transferible a nuevos contextos.

Al mismo tiempo, se hace patente una desigualdad geográfica de la investigación: dos tercios de los estudios provienen de Europa y Norteamérica, mientras que Latinoamérica aporta menos del 15 %, con apenas un caso ecuatoriano. Este desequilibrio sugiere que las soluciones importadas podrían no ajustarse a los marcos culturales, económicos y regulatorios locales. Por consiguiente, las universidades y centros de investigación latinoamericanos disponen de una oportunidad estratégica para desarrollar prototipos pedagógicos y tecnológicos que integren la alfabetización algorítmica con los contextos sociopolíticos de la región.

En el plano teórico, el trabajo amplía el modelo de resiliencia cognitiva: a los pilares clásicos de pensamiento crítico e inoculación discursiva se suma la dimensión algorítmica, que traslada parte de la responsabilidad de la veracidad a la arquitectura misma de los sistemas de recomendación. Este cambio de foco implica que la alfabetización mediática ya no puede limitarse a ‘leer y verificar’, sino que debe enseñar a ‘interrogar’ los entornos digitales: por qué vemos lo que vemos, quién decide la jerarquía de los contenidos y qué sesgos arrastra el diseño de ese flujo.

La síntesis también arroja alertas. Primero, existe el riesgo de que la alfabetización mediática se convierta en una ‘solución parche’ si no va acompañada de políticas de transparencia algorítmica y rendición de cuentas empresarial. Por más competente que sea un ciudadano, sus posibilidades

de discernir se reducen cuando la interfaz oculta las señales de procedencia y manipula métricas de popularidad. Segundo, la excesiva dependencia de modelos de IA propiedad de grandes corporaciones puede reproducir los sesgos que la alfabetización intenta combatir. Las iniciativas de código abierto y la auditoría cívica deben considerarse parte del ecosistema de alfabetización.

Los beneficios de invertir en alfabetización mediática integrada, sin embargo, son palpables. A corto plazo, las universidades que incorporan módulos de detección de *deepfakes* o talleres de transparencia algorítmica registran incrementos en la autoconsciencia digital y en la responsabilidad de compartir información. A mediano plazo, los laboratorios de *factchecking* con IA abierta pueden servir de puentes entre medios públicos, investigadores y ciudadanía, acelerando la respuesta a bulos localmente relevantes. A largo plazo, formar profesionales capaces de auditar algoritmos y evaluar narrativas sintéticas contribuirá a consolidar democracias digitales más robustas.

Recomendaciones para la práctica educativa y la política pública:

1. Incluir en los primeros semestres de las carreras de Comunicación, Periodismo e Informática una asignatura obligatoria de educación algorítmica y detección de medios sintéticos.
2. Establecer laboratorios de verificación con IA de código abierto en los medios públicos y universidades, de modo que alumnos, periodistas y programadores colaboren en proyectos reales de desmentido.
3. Capacitar a docentes de todos los niveles en metodologías de prebunking y lectura lateral, dotándolos de guías didácticas que integren ejercicios con *chatbots* y detectores de *deepfakes*.
4. Impulsar políticas de transparencia algorítmica, obligando a las plataformas a documentar sus sistemas de recomendación y a ofrecer paneles de control comprensibles para los usuarios.
5. Priorizar campañas de alfabetización mediática en poblaciones vulnerables—adultos mayores, comunidades rurales, estudiantes con baja conectividad—donde la exposición a bulos es alta y la capacidad de verificación, limitada.

Líneas futuras de investigación

Se necesitan estudios longitudinales que midan la persistencia de la resiliencia informativa y exploren variables mediadoras (motivación política, autoeficacia digital). Igualmente, es urgente experimentar con interfaces pedagógicas basadas en realidad aumentada o simulaciones inmersivas que permitan ‘ver’ el recorrido de los datos a través de los algoritmos. Por último, convendría desarrollar índices que midan la calidad de la alfabetización algorítmica como competencia ciudadana, distinguiendo entre comprensión declarativa y uso efectivo.

Los hallazgos sintetizados no agotan el tema, pero delinean un camino factible: combinar saber crítico y saber técnico para neutralizar un fenómeno que evoluciona a la velocidad del cómputo. La alfabetización mediática, reforzada por la inteligencia artificial y arropada por políticas de transparencia, puede convertirse en un cortafuegos democrático capaz de frenar la propagación de la desinformación automatizada sin sacrificar la apertura informativa que caracteriza a la era digital.

Referencias

- Aparici, R., Bordignon, F. R. A., y Martínez-Pérez, J. (2021). Alfabetización algorítmica basada en la metodología de Paulo Freire. *Perfiles Educativos*, 43(Especial), 36–54. <https://doi.org/10.22201/iisue.24486167e.2021.Especial.61019>
- Barrientos-Báez, A., Piñeiro Otero, M. T., y Porto Renó, D. (2024). Imágenes falsas, efectos reales. Deepfakes como manifestaciones de la violencia política de género. *Revista Latina de Comunicación Social*, 82, 1–30. <https://doi.org/10.4185/rllcs-2024-2278>
- Baumann, F., Arora, N., Rahwan, I., y Czaplicka, A. (2025). *Dynamics of Algorithmic Content Amplification on TikTok* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2503.20231>
- Braun, V., y Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Brodsky, J. E., Brooks, P. J., Scimeca, D., Galati, P., Todorova, R., y Caulfield, M. (2021). Associations Between Online Instruction in Lateral Reading Strategies and Fact-Checking COVID-19 News Among College Students. *AERA Open*, 7, 23328584211038937. <https://doi.org/10.1177/23328584211038937>
- Brodsky, J. E., Brooks, P. J., Scimeca, D., Todorova, R., Galati, P., Batson, M., Grosso, R., Matthews, M., Miller, V., y Caulfield, M. (2021a). Improving college students' fact-checking strategies through lateral reading instruction in a general education civics course. *Cognitive Research: Principles and Implications*, 6(1), 23. <https://doi.org/10.1186/s41235-021-00291-4>
- Bulger, M., y Davison, P. (2018). The Promises, Challenges, and Futures of Media Literacy. *Journal of Media Literacy Education*, 10(1), 1–21. <https://doi.org/10.23860/JMLE-2018-10-1-1>
- Cheung, J. C., y Ho, S. S. (2025). Explainable AI and trust: How news media shapes public support for AI-powered autonomous passenger drones. *Public Understanding of Science*, 34(3), 344–362. <https://doi.org/10.1177/09636625241291192>
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., y Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>
- Council of Europe. (2024). *IRIS Special: Focus on media literacy*. European Audiovisual Observatory. <https://rm.coe.int/iris-2024-2-media-literacy/1680b06196>
- Dogruel, L., Masur, P., y Joeckel, S. (2022). Development and Validation of an Algorithm Literacy Scale for Internet Users. *Communication Methods and Measures*, 16(2), 115–133. <https://doi.org/10.1080/19312458.2021.1968361>
- Ferrara, E., Cresci, S., y Luceri, L. (2020). Misinformation, manipulation, and abuse on social media in the era of COVID-19. *Journal of Computational Social Science*, 3(2), 271–277. <https://doi.org/10.1007/s42001-020-00094-5>
- Fieiras Ceide, C., Vaz Álvarez, M., y Túñez López, M. (2022). Verificación automatizada de contenidos en las radiotelevisións públicas

- europas: Primeras aproximaciones al uso de la inteligencia artificial. *Redmarka. Revista de Marketing Aplicado*, 26(1), 36–51. <https://doi.org/10.17979/redma.2022.26.1.8932>
- Gagrčin, E., Naab, T. K., y Grub, M. F. (2024). Algorithmic media use and algorithm literacy: An integrative literature review. *New Media & Society*, 14614448241291137. <https://doi.org/10.1177/14614448241291137>
- Guo, Z., Schlichtkrull, M., y 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
- Haddaway, N. R., Page, M. J., Pritchard, C. C., y McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*, 18(2), e1230. <https://doi.org/10.1002/cl2.1230>
- Herrero-Diz, P., Pérez-Escolar, M., y Varona Aramburu, D. (2022). Competencias de verificación de contenidos: Una propuesta para los estudios de Comunicación. *Revista de Comunicación*, 21(1), 231–249. <https://doi.org/10.26441/RC21.1-2022-A12>
- Jones-Jang, S. M., Mortensen, T., y Liu, J. (2021). Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't. *American Behavioral Scientist*, 65(2), 371–388. <https://doi.org/10.1177/0002764219869406>
- Kong, S.-C., Man-Yin Cheung, W., y Zhang, G. (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers and Education: Artificial Intelligence*, 2, 100026. <https://doi.org/10.1016/j.caeai.2021.100026>
- Landis, J. R., y Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Lewandowsky, S., y Van Der Linden, S. (2021). Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology*, 32(2), 348–384. <https://doi.org/10.1080/10463283.2021.1876983>
- Lim, G., y Perrault, S. T. (2023). XAI in Automated Fact-Checking? The Benefits Are Modest and There's No One-Explanation-Fits-All. *Proceedings of the 35th Australian Computer-Human Interaction Conference*, 624–638. <https://doi.org/10.1145/3638380.3638388>
- Maertens, R., Roozenbeek, J., Basol, M., y Van Der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1–16. <https://doi.org/10.1037/xap0000315>
- Montoya Ramírez, N., Zuluaga Arias, L., y Rivera-Rogel, D. (2020). Periodismo y competencias mediáticas: Una aproximación desde contexto colombiano y ecuatoriano. *Revista de Comunicación*, 19(1), 185–206. <https://doi.org/10.26441/RC19.1-2020-A11>
- Noguera-Vivo, J. M., y Grandío-Pérez, M. D. M. (2025). Enhancing Algorithmic Literacy: Experimental Study on Communication Students' Awareness of Algorithm-Dri-

- ven News. *Anàlisi*, 71, 37–53. <https://doi.org/10.5565/rev/analisi.3718>
- Oeldorf-Hirsch, A., y Neubaum, G. (2025). What do we know about algorithmic literacy? The status quo and a research agenda for a growing field. *New Media & Society*, 27(2), 681–701. <https://doi.org/10.1177/14614448231182662>
- Quevedo-Redondo, R., Gómez-García, S., y Navarro-Sierra, N. (2022). Aprendiendo a desinformar: Una estrategia de prebunking con newsgames para estimular la adquisición de competencias en el grado en Periodismo. *Anàlisi*, 66, 95–112. <https://doi.org/10.5565/rev/analisi.3447>
- Ramírez-Montoya, M.-S., y Lugo-Ocando, J. (2020). Systematic review of mixed methods in the framework of educational innovation. *Comunicar*, 28(65), 9–20. <https://doi.org/10.3916/C65-2020-01>
- Sádaba, C., y Salaverría, R. (2022). Combatir la desinformación con alfabetización mediática: Análisis de las tendencias en la Unión Europea. *Revista Latina de Comunicación Social*, 81, 17–33. <https://doi.org/10.4185/RLCS-2023-1552>
- Scheibenzuber, C., Hofer, S., y Nistor, N. (2021). Designing for fake news literacy training: A problem-based undergraduate online-course. *Computers in Human Behavior*, 121, 106796. <https://doi.org/10.1016/j.chb.2021.106796>
- Suarez-Lledo, V., y Alvarez-Galvez, J. (2022). Assessing the Role of Social Bots During the COVID-19 Pandemic: Infodemic, Disagreement, and Criticism. *Journal of Medical Internet Research*, 24(8), e36085. <https://doi.org/10.2196/36085>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- UNESCO. (2021). *Normas mundiales para el desarrollo de planes de estudios de alfabetización mediática e informacional: Directrices*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. https://www.unesco.org/sites/default/files/medias/files/2022/02/Global%20Standards%20for%20Media%20and%20Information%20Literacy%20Curricula%20Development%20Guidelines_ES.pdf
- Vaccari, C., y 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), 2056305120903408. <https://doi.org/10.1177/2056305120903408>
- Vizoso, Á., y Vázquez-Herrero, J. (1970). Fact-checking platforms in Spanish. Features, organisation and method. *Communication & Society*, 127–143. <https://doi.org/10.15581/003.32.37819>
- Wedlake, S., Coward, C., y Lee, J. H. (2024). How games can support misinformation education: A sociocultural perspective. *Journal*

of the Association for Information Science and Technology, 75(13), 1480–1497. <https://doi.org/10.1002/asi.24954>

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11), 39–52. <https://doi.org/10.22215/timreview/1282>

World Health Organization. (2020). *Ethical considerations to guide the use of digital proximity tracking technologies for COVID-19 contact tracing*. World Health Organization. <https://iris.who.int/bitstream/handle/10665/334287/9789240010314-eng.pdf>

Zhang, H., Zhu, Z., y Caverlee, J. (2023). *Evolution of Filter Bubbles and Polarization in News Recommendation* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2301.10926>