

Combatiendo la desinformación: una revisión crítica de la literatura

Fighting disinformation: A critical literature review

CARLOS MUNOZ-SERNA¹
ALEJANDRO MURILLO-LARA²

Resumen

La proliferación de desinformación se presenta como una amenaza para los procesos democráticos, la salud pública, la estabilidad de las instituciones de la sociedad civil, y la seguridad de la población. La percepción de que esta amenaza es cada vez mayor ha hecho que una multitud de organizaciones de la sociedad civil, gobiernos, redes sociales e investigadores de las más diversas áreas hayan puesto su atención en el diseño e implementación de estrategias que permitan limitar los efectos nocivos de la desinformación. A partir de un análisis crítico de la literatura científica existente, este artículo identifica las estrategias propuestas para hacer frente a lo que algunos han llamado como el fenómeno del desorden informativo, y hace una evaluación de los obstáculos más importantes que ellas enfrentan.

Palabras clave: desorden informativo, desinformación, educación digital, fact-checking, prevención cognitiva.

Abstract

The proliferation of disinformation poses a threat to democratic processes, public health, the stability of civil society institutions, and the safety of the population. The perception that this threat is growing has led a wide range of civil society organizations, governments, social networks, and researchers across fields to focus on designing and implementing strategies to limit the harmful effects of disinformation. Based on a critical analysis of the existing scientific literature, this article identifies the strategies

¹ University of Cincinnati, munozca@mail.uc.edu, <https://orcid.org/0000-0002-0092-2086>

² Universidad Pedagógica y Tecnológica de Colombia, luisalejandro.murillo@uptc.edu.co, <https://orcid.org/0000-0002-8156-4128>

proposed to address what some have called the phenomenon of information disorder and evaluates the most significant obstacles they face.

Key words: information disorder, disinformation, digital education, fact-checking, cognitive prevention.

1. Introducción

En la última década, la desinformación y sus efectos en la esfera pública ha sido un tema que causa cada vez mayor preocupación (Newman et al. 2020; Orth y Bialik 2025). Tanto ha crecido el volumen de desinformación producida, y tanto han aumentado sus efectos negativos, que algunos autores consideran que debemos señalar un punto de ruptura. Así, algunos autores consideran que hemos entrado en una época caracteriza por una crisis epistémica (McIntyre 2023), mientras que otros aducen que estamos acercándonos a una nueva era, la de la post-verdad. Según estos autores, la era de la post-verdad se caracteriza por el hecho de que hay una desconfianza generalizada en el conocimiento experto, el conocimiento científico se considera “elitista”, el criterio y opiniones de los expertos están a la par de cualquiera, y el consenso de la comunidad científica ha perdido peso a la hora de determinar lo que debemos considerar como verdadero en la esfera pública (Lewandowsky, Ecker y Cook 2017).

Los peligros asociados a la desinformación no son menores. La proliferación de desinformación sea esta diseminada de manera intencional o no, se ha asociado con la manipulación política y electoral, con el fomento de la polarización y el debilitamiento de la esfera pública, y la erosión de la confianza en las instituciones. Estos procesos, en su conjunto, se configuran como una amenaza para la construcción de democracias sólidas. Como lo propone Chris Tenove (2020: 518), la desinformación pone en jaque tres de los valores sobre los que se fundan los sistemas democráticos: la autodeterminación ciudadana de las políticas implementadas en un Estado; la representación política que es premiada o reprendida a través de elecciones justas; y la deliberación pública para formar opiniones y voluntades.

Pero el aumento de desinformación no solo ha sido un tema de preocupación por cuestiones políticas, sino que también lo ha sido porque ella tiene el potencial para poner en riesgo la salud y bienestar de la población. Así, por ejemplo, son muchas las ocasiones en las que la desinformación ha sido asociada con actos violentos. Entre los muchos ejemplos en los que la desinformación se ha asociado con actos violentos destacan el caso de Pizzagate en Washington y la proliferación de teorías conspirativas entre la extrema derecha estadounidense; el genocidio de la comunidad Rohingya en Myanmar, en el que Facebook sirvió de plataforma para la diseminación de mensajes de odio (Schissler 2025); o el incidente de Torre Pacheco, España, en el que información falsa sobre un supuesto ataque por parte de un ciudadano de ascendencia marroquí motivó una oleada de violencia xenófoba en contra de la población norteafricana.

Los potenciales peligros de la desinformación fueron aún más patentes durante la pandemia de Covid-19. En este punto hubo tal abundancia de información falsa y engañosa que la Organización Mundial de la Salud, en conjunto con otras agencias internacionales, se vio obligada a alertar que, a la par que estábamos en medio una pandemia viral, estábamos frente a una *infodemia* (World Health Organization 2020). La falta de confianza en los discursos científicos avalados institucionalmente, o el aumento desmesurado de información falsa sobre cómo afrontar la emergencia, hizo que las personas adoptaran comportamientos que aumentaba su riesgo de contagio, o que afectaba negativamente su salud. Así, la desinformación estuvo asociada adopción de medidas y tratamientos que o bien no eran efectivos, en el mejor de los casos, o que eran directamente perjudiciales, en el peor.

Los peligros que supone la desinformación han hecho que se produzca un aumento pronunciado y continuado de trabajos que analizan este fenómeno. Entre los múltiples acercamientos que se pueden encontrar destacan, en primer lugar, aquellos que analizan qué es la desinformación y cómo se relaciona con otros conceptos que le son afines como los de *fake-news*, *misinformation*, *malinformation*, *disinformation*, *clickbait*, sátira, mentir, entre otros. En segundo, destacan los trabajos que analizan los factores psicológicos, sociológicos, lingüísticos, políticos y técnicos que causan o intensifican este fenómeno. El tercer acercamiento que se debe destacar, y el que va a ser objeto de análisis en este trabajo, analiza las estrategias que se han implementado o se pueden implementar para hacer frente a la avalancha de desinformación. Profesionales de múltiples ciencias, desde periodistas hasta psicólogos, pasando por programadores y educadores, han aunado esfuerzos para pensar y poner en práctica medidas que disminuyan la información falsa o engañosa que corre en redes, se limiten las cámaras de eco en las que muchos nos encontramos, se disminuya el efecto nocivo de este tipo de mensajes, y se fortalezcan las capacidades de los usuarios para navegar de manera más segura el ecosistema informativo.

En este trabajo se pone la lupa sobre las diferentes aproximaciones que se han propuesto para combatir la desinformación. Así, en este artículo se investiga cuáles son las más importantes tendencias a la hora de intentar combatir la desinformación, al tiempo que se desarrolla un ejercicio crítico de dichas propuestas. En la primera parte de este trabajo, se realiza una revisión de la literatura en la que se identifican, categorizan y analizan 39 artículos que estudian estrategias para combatir la desinformación. A partir de este trabajo, se desarrolla un mapa de las diferentes estrategias encontradas, las cuales son organizadas en 5 categorías (a. Verificación y Análisis de Contenido b. Educación y Alfabetización c. Prevención Cognitiva d. Comunicación y Contranarrativa e. Intervenciones Sistémicas). Esta forma de organizar las decenas de estrategias identificadas permite ver la continuidad entre ellas y entender que, a pesar de que en la literatura pueda haber multiplicidad de acercamientos al problema de la desinformación, en el fondo, los acercamientos tienen mucho en común. Finalmente, se lleva a cabo un ejercicio evaluativo de las estrategias en el que se analizan algunos de los obstáculos clave que pueden enfrentar dichas propuestas.

2. Desinformación: un concepto polisémico

En las discusiones acerca de la producción y diseminación de información dudosa, falsa, o malintencionada en los espacios informativos hay una multitud de conceptos que, a pesar de tener similitudes, deben ser distinguidos para aumentar la precisión de la discusión. Probablemente, los tres más importantes son *mis-información*, *des-información*, y *mal-información*.³ Estos tres conceptos, que han sido retomados de las discusiones en habla inglesa, se diferencian en relación con la veracidad del mensaje que es difundido, así como con la intención de los emisores (Wardle y Derakhshan 2017; Sádaba y Salaverría 2023).

a) Des-información: se caracteriza por la diseminación de información que se sabe falsa, con la que se pretende engañar al público, para así causar algún tipo de daño a una persona, un grupo social, o una institución.

b) Mis-información: este fenómeno se caracteriza por la diseminación accidental de información falsa, que además se produce sin la intención de causar daño o de manipular maliciosamente la opinión pública.

c) Mal-información: diseminación de mensajes que son verdaderos, o parcialmente verdaderos, que se difunden con la intención de causar algún tipo de daño o manipular la opinión pública.

Los tres fenómenos mencionados configuran lo que se ha llamado como el desorden informativo. Este fenómeno se refiere a la circulación de información con un bajo nivel de veracidad, comprometida por la existencia de intereses políticos o económicos (Irwanto et al. 2025; Wardle y Derakhshan 2017).

Otros conceptos han cobrado relevancia en la discusión sobre los múltiples tipos de desorden informativo, sin embargo, se considera que los anteriores tres conceptos (des- mis- mal- información) son los centrales, en tanto que ellos subsumen los otros conceptos que pueden encontrarse en la literatura. Por ejemplo, un término que se ha vuelto popular en estas discusiones es el de *fake-news*, el cual se define como un tipo de información falsa, que es difundida de manera intencional por un medio de noticias (Zhou y Zafarani 2019). De este modo, es posible conceptualizar las *fake-news* como un tipo de desinformación, que se diferencia por el actor produce la información. Este hecho, en conjunción con el uso abusivo que algunos actores en el ecosistema informativo han hecho del concepto de *fake-news* para deslegitimar el ejercicio periodístico, ha llevado a que muchos autores prefieran acudir a conceptos alternativos para analizar estos fenómenos (Sádaba y Salaverría 2023). Este es el caso del presente trabajo, en el que se hará referencia al fenómeno del desorden informativo de manera general para recoger todos los otros tipos fenómenos que se asocian con la contaminación del ecosistema informativo.

³ Estos conceptos se corresponden con los vocablos en lengua inglesa de *misinformation*, *malinformation*, y *disinformation*, lengua en la que se produce el grueso de las discusiones sobre el fenómeno desinformativo que se está tratando.

En este trabajo se contemplan estrategias que pretenden hacer frente a los tres tipos de fenómenos asociados con el desorden informativo. A primera vista, algunas de las estrategias que se encontraron están pensadas para hacer frente de manera específica a la des-información y la mis-información. Tal podría ser el caso *fact-checking*, el cual se centra en determinar la veracidad de la información producida. Sin embargo, el fact-checking, en ocasiones también hace esfuerzos para combatir la mal-información. Así, algunos agentes verificadores no solo reportan que determinada información es falsa o verdadera, sino que muestran que cierta información es parcialmente falsa, engañosa, o que necesita mayor contexto para poder ser comprendida. Más aún, en este trabajo se identificaron estrategias que pretenden crear consumidores de información con un mayor sentido crítico que son capaces de juzgar de mejor manera la información ante la que se encuentran. De este modo, por ejemplo, las estrategias que se centran en la alfabetización digital promueven la construcción de una serie de habilidades que les permitan a los usuarios hacer frente a la des-información, la mal-información y la mis-información.

3. Metodología

En este trabajo pretende responder dos preguntas de investigación, una de ellas descriptiva y la otra evaluativa.

Pregunta de investigación 1. ¿Cuáles son las estrategias que se han planteado para combatir la abundancia de información falsa o engañosa en el ecosistema de medios comunicativos?

Pregunta de investigación 2. ¿Cuáles son los obstáculos que enfrentan las estrategias que son más populares en la literatura para ser exitosas a la hora de reducir la cantidad como los efectos negativos de la información falsa o engañosa?

Para responder a estas dos preguntas de investigación se realizó una revisión crítica de la literatura. La revisión crítica de la literatura se caracteriza por hacer una síntesis de la investigación existente sobre un tema, para posteriormente hacer un ejercicio evaluativo que a su vez determine las fortalezas, debilidades y vacíos en la literatura analizada (Atlas.ti 2025; Grant y Booth 2009: 93). Siguiendo este esquema, en el presente trabajo se pueden identificar dos pasos claros. Por un lado, se encuentra la parte descriptiva, en la que se realiza una reconstrucción del panorama investigativo sobre las diferentes medidas que se podrían implementar para hacer frente al fenómeno del desorden informativo. Por el otro, se encuentra una parte crítica, en la que se lleva a cabo un balance de algunas de las tendencias más importantes en la literatura.

Para poder completar el primer paso de esta investigación, el ejercicio descriptivo, se ha desarrollado una revisión de la literatura, en la que se identifican, clasifican y describen las investigaciones más recientes y destacadas en el campo investigativo sobre cómo luchar contra el desorden informativo. De manera específica, esta revisión siguió los pasos que enumeramos a continuación:

I. Selección de bases de datos: se seleccionaron cuatro bases de datos para realizar la búsqueda de archivos, a saber, *EBSCO: Communication & Mass Media Complete*, *PsycINFO*, *JSTOR*, y *Scopus* (Elsevier).

El fenómeno desinformativo debe ser analizado y enfrentado de manera interdisciplinaria. Por esta razón se eligieron bases de datos que contienen trabajos de las más diversas disciplinas. En primer lugar, se seleccionó *EBSCO: Communication & Mass Media Complete* debido a que es una de las bases de datos más completas en relación con estudios sobre medios de comunicación y periodismo. Segundo, se eligió incluir la base de datos *PsycINFO* debido a que esta es la base de datos más relevante sobre trabajos producidos en los campos de las ciencias cognitivas y la psicología. En tercer lugar, se eligió hacer una búsqueda en *JSTOR*, en tanto que esta proporciona una mirada más amplia del fenómeno de la desinformación desde las humanidades y las ciencias sociales. Finalmente, se utilizó *Scopus*, en tanto que esta es una de las bases de datos multidisciplinaria más amplias de literatura revisada por pares. En su conjunto, estas cuatro bases de datos facilitan una mirada comprensiva sobre los estudios que analizan las estrategias para hacer frente al desorden informativo.

II. Criterios de búsqueda: la búsqueda en las cuatro bases de datos anteriormente identificadas incluyó tres criterios.

a) Periodo de revisión: artículos publicados entre 2015 y 2025⁴. Se eligió enfocarse en este periodo para dar una perspectiva actualizada de este fenómeno. El fenómeno del desorden informativo es extremadamente dinámico y la preocupación por sus efectos ha ido en aumento durante la última década (Navarro-Sierra, Magro-Vela y Vinader-Segura 2024).

b) Tipos de documentos: artículos científicos en español o inglés, revisados por pares. Se eligió incluir solo artículos científicos revisados por pares para asegurar que los documentos revisados cumplen con altos estándares de rigurosidad y calidad. Se incluyeron artículos en inglés en tanto que esta es la lengua franca de la producción científica. Más del 90% de los artículos producidos a este respecto son escritos en idioma inglés (Navarro-Sierra, Magro-Vela y Vinader-Segura 2024). Aunque la producción en idioma español es comparablemente menor a la que se puede encontrar en otras lenguas (Navarro-Sierra, Magro-Vela y Vinader-Segura 2024), se decidió incluir en la búsqueda artículos en este idioma para también reflejar las particularidades del contexto hispanoamericano. Este último enfrenta circunstancias sociales, económicas y políticas particulares, y cuenta con una serie de mecanismos institucionales relacionados con el fenómeno del desorden informativo que se diferencian de los que podemos encontrar, por ejemplo, en el mundo anglosajón (Herrero y Herrera Damas 2021; Moreno-Gil, Ramon y Rodríguez-Martínez 2021).

⁴ En esta selección se realizó una excepción al incluir el artículo titulado *Misinformation and Its Correction: Continued Influence and Successful Debiasing*, el cual fue publicado en el año 2012. Esta excepción fue motivada por dos razones. En primer lugar, este artículo es citado por varios de los artículos que fueron incluidos en el canon final que fue analizado en este trabajo. En segundo, los trabajos de Lewandowsky y su equipo han tenido una influencia decisiva en la investigación sobre la naturaleza de la desinformación, sus efectos y las estrategias para hacerle frente.

c) Palabras clave: se siguió el método de búsqueda booleano, en el que se buscaron en las bases de datos seleccionadas las palabras clave detalladas en la tabla 1. Estas palabras fueron buscadas en español y en inglés para capturar bibliografía producida en Hispanoamérica o que se enfocase en analizar esta región. La combinación de palabras clave incluidas en esta búsqueda permitió no solamente identificar los trabajos relacionados con el fenómeno del desorden informativo, sus causas o sus consecuencias, sino aquellos que específicamente centraban su atención en analizar estrategias para combatirlo.

Tabla 1. Lista de palabras clave y operadores booleanos usados para buscar artículos relevantes

(Disinformation OR misinformation OR fake news OR cognitive biases OR manipulation)
AND
(Media OR social media)
AND
(Strategies against disinformation OR digital education OR content curation OR data verification OR fact checking OR media literacy OR Countering disinformation OR fight disinformation)

Fuente: Elaboración propia.

III. Preselección de artículos: Después de buscar en las bases de datos mencionadas, se ordenaron los artículos por relevancia y por el número de citas con las que contaba, lo que permitió filtrar aquellos que eran más importantes dado el tema analizado. Después de realizar esta búsqueda, se obtuvieron 148 referencias que cumplían con los criterios de búsqueda establecidos para esta revisión.

IV. Selección de artículos: Después de haber identificado un conjunto de 148 artículos que cumplía con los criterios de búsqueda establecidos, se procedió a leer los resúmenes de cada uno de ellos para determinar cuáles tenían como objetivo principal analizar o proponer estrategias contra la desinformación. Para esta selección, los dos investigadores del proyecto leyeron todos los resúmenes y de manera independiente determinaron cuáles debían ser incluidos y cuáles no. Todos los artículos seleccionados por ambos investigadores pasaron a formar parte del corpus de análisis, mientras que aquellos en los que no hubo consenso inicial fueron posteriormente discutidos entre los investigadores para determinar si tenían valor analítico a la hora de responder las preguntas de investigación.

De manera específica, los artículos que fueron seleccionados fueron aquellos cuyo objetivo central era analizar estrategias para hacerle frente al fenómeno del desorden informativo. Mientras tanto, aquellos artículos que se centraban en analizar las causas, las consecuencias o las variables asociadas al fenómeno del desorden informativo fueron dejados de lado.

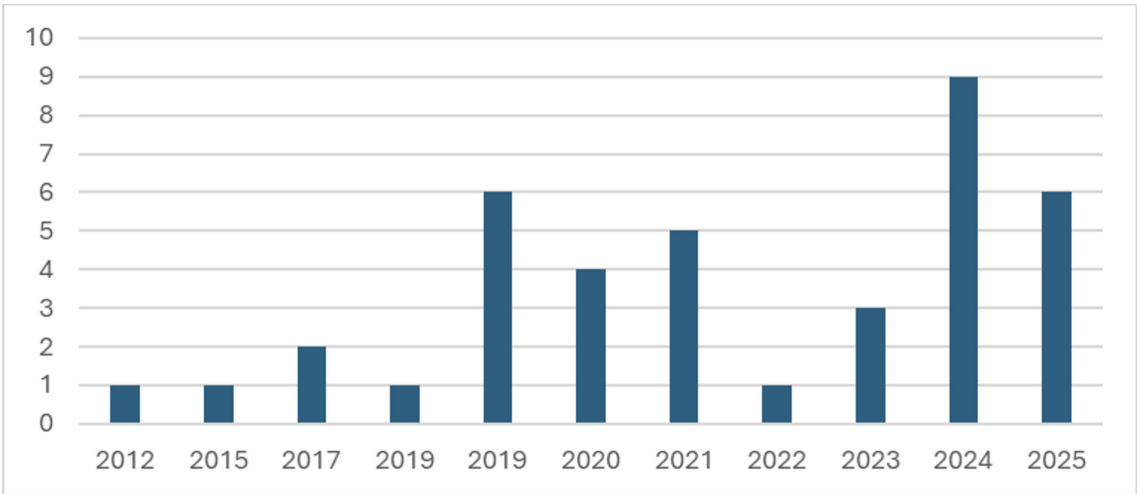
El último criterio usado para seleccionar los artículos que hicieron parte del canon final es su índice de citación. Considerar el índice de citación a la hora de elegir los artículos que harían parte del corpus de análisis final permitió que estos artículos fueran representativos de las tendencias en el tema analizado, y que la discusión y el análisis hecho en este trabajo fuese sobre artículos con alto impacto, académicamente relevantes, y que se han consolidado en las discusiones sobre el fenómeno de la desinformación y las estrategias para hacerle frente. Así, a partir de la lista inicial de 148 artículos que se identificaron inicialmente en las bases de datos, se conformó un corpus de análisis de 39 artículos.

V. Análisis y sistematización de los documentos: los investigadores leyeron los artículos seleccionados y construyeron una ficha RAE por cada uno de ellos. En estas fichas se describió el documento, se detalló la metodología usada, se indicaron las conclusiones del artículo, se identificaron las estrategias que eran objeto de análisis, se revisaron las referencias usadas, y se detalló el aporte que el artículo hacía a los estudios sobre comunicación.

4. Resultados

Se revisó un total de 39 artículos agrupados en tres tipologías: investigación (n=22), revisión (n=7) y reflexión (n=10). Todos los artículos fueron publicados entre 2015 y 2025, con excepción de uno publicado en 2012. De los 39 artículos analizados, 18 fueron publicados a partir de 2023; 10 artículos fueron publicados entre 2020 y 2022; y 10 artículos fueron publicados entre 2015 y 2019. Como puede observarse, 28 de los 39 artículos revisados fue publicado de 2020 en adelante (Imagen 1).

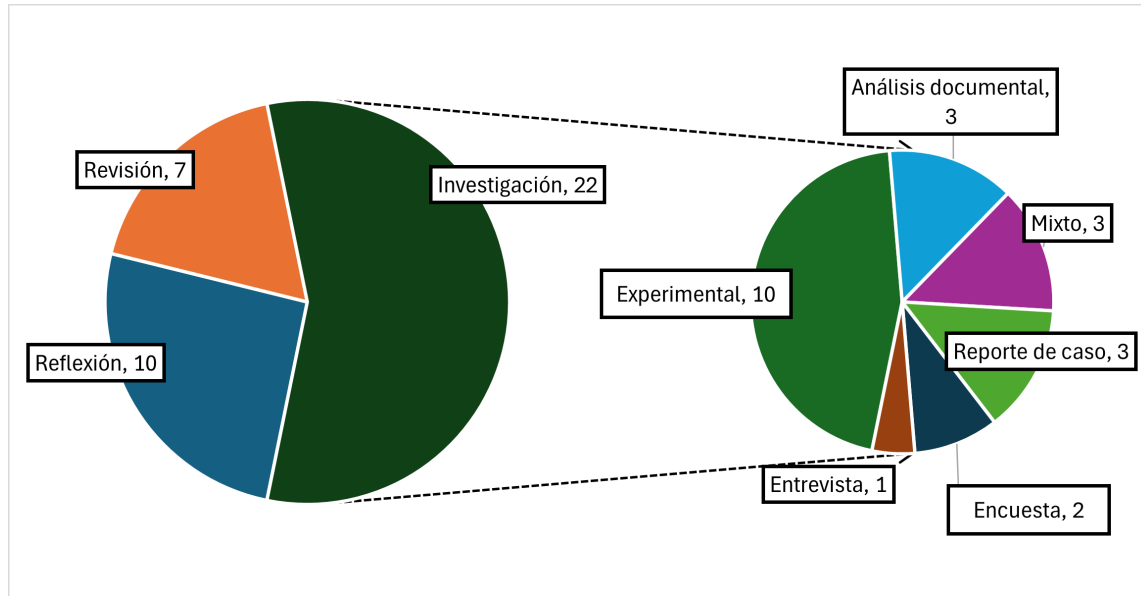
Imagen 1. Número de artículos publicados por año



Fuente: Elaboración propia.

Los artículos de investigación, el subgrupo de artículos más amplio, siguieron metodologías diversas, tales como experimental (n=10), análisis documental (n= 3), reporte de caso (n=3), encuesta (n=2), entrevistas (n=1), y mixto (n=3). Mientras tanto, entre los artículos de revisión se encontraron revisiones sistemáticas (n=5) y metaanálisis (n=2) (Imagen 2).

Imagen 2. Número de artículos por tipo y metodología usada



Fuente: Elaboración propia.

El análisis del contenido de los artículos revisados se enfocó en las estrategias para combatir la desinformación consideradas en cada uno. A partir de dicho análisis, se elaboró una lista de 19 estrategias (Tabla 2).

4.1. Análisis de las estrategias

Al analizar las estrategias, se encontró que estas compartían rasgos comunes que permitió organizarlas en 5 grupos, estos son: a. Verificación y Análisis de Contenido, b. Educación y Alfabetización, c. Prevención Cognitiva, d. Comunicación y Contranarrativa, y e. Intervenciones Sistémicas. En la tabla 2 están referenciados los artículos en los que se discute cada estrategia. Vale la pena notar que en esta tabla se puede observar que hay artículos que aparecen asociados con más de una estrategia, esto se debe a que, como muchos de los artículos mismos reconocen, la lucha contra la desinformación no puede hacerse desde un solo frente.

4.1.1. Verificación y Análisis de Contenido

Se trata de estrategias reactivas que evalúan la veracidad de contenidos ya existentes mediante verificación de hechos, análisis de forma y sistemas de calificación.

a) Accuracy nudging: Intervención que tiene como objetivo interrumpir el modo de procesamiento de información por defecto de los individuos, que los lleva a asumir que la información que encuentran es veraz. Esta estrategia consiste en usar mensajes simples para que las audiencias tomen una actitud más crítica frente a la información y se pregunten si la información que están consumiendo es veraz o no.

b) Estrategias implementadas por la comunidad: Estrategias en las que es la comunidad de usuarios quien vigila o cura el contenido encontrado en redes. Esto se puede hacer por medio de la publicación de mensajes correctivos, o por medio de estrategias menos demandantes, como puede ser la publicación de tags. En particular, destaca la estrategia de los cuestionamientos a la veracidad, en la que los usuarios hacen preguntas que cuestionan los criterios usados para validar una información como certera.

c) Fact-checking: Esta estrategia consiste en hacer una verificación de la información publicada, ya sea en redes, o en medios tradicionales. Adicionalmente, se destaca que en esta investigación la verificación fue analizada en relación con los esfuerzos conducidos por organismos especializados en esta labor. Sin embargo, en algunas instancias el fact-checking puede ser realizado por instituciones que, a pesar de que no se especializan en la verificación de datos, tienen conocimiento experto sobre temas específicos y realizan chequeo sobre información sobre dichos temas. Un caso representativo puede ser el del Instituto Nacional de Alergias y Enfermedades Infecciosas (NIAID), el cual, en cabeza de Anthony Fauci, realizó chequeo de información durante la pandemia de COVID-19.

Dada la prevalencia que tuvo el fact-checking en la presente revisión de la literatura, es preciso destacar algunos de los subtipos asociados a esta estrategia.

Automático o Manual: El fact-checking manual es aquel que llevan a cabo personas expertas. Ellas identifican y corrigen los mensajes falsos o engañosos. En contrapartida, el fact-checking automático supone la introducción de inteligencias artificiales para hacer la detección y en ocasiones la corrección de la información. Sin embargo, parece haber un consenso sobre la necesidad de adoptar un enfoque híbrido, en el que las inteligencias artificiales son herramientas que potencian las capacidades de los verificadores, pero no los reemplazan.

Contenido, Estilo, Propagación, Credibilidad de la fuente: Uno de los problemas más importantes que enfrentan los verificadores de contenidos versa sobre cómo detectar las noticias falsas. Entre los múltiples métodos para hacerlo, destacan cuatro; (1) Atención al contenido. Es la manera más directa y clara de verificación, ella consiste en verificar si la información concuerda con el conocimiento

preexistente; (2) Atención al estilo. Este método consiste en observar los patrones usados para la construcción del mensaje. Así, por ejemplo, uno de los criterios para detectar noticias falsas es si esta está construida de tal manera que induce respuestas emocionales fuertes. Este patrón está asociado con la producción de desinformación; (3) Atención a los patrones de propagación. Este método observa la manera en que se difunde el mensaje, para detectar si sigue el comportamiento de mensajes verdaderos o falsos; y (4) Atención a la credibilidad de la fuente. Este método consiste en investigar la credibilidad de las fuentes que publican o reproducen la información.

4.1.2. Educación y Alfabetización

Este grupo de estrategias, también conocidas con otros nombres como alfabetización informacional, digital, mediática o educomunicación; reúne los enfoques orientados a desarrollar en las audiencias mejores habilidades de pensamiento crítico que les permitan acceder, analizar, y evaluar información en diferentes contextos y plataformas; así como para producir y reproducir información. Estas estrategias buscan formar ciudadanos más preparados para identificar y resistir la desinformación, la mis-información y la mal-información de manera autónoma.

- a) Formación práctica: Este enfoque busca desarrollar las habilidades críticas de los espectadores por medio de su exposición a situaciones en las que tienen que interactuar y juzgar información que puede ser falsa, dudosa o desorientadora.
- b) Guías: Este tipo de intervenciones implican proveer a las audiencias con documentos que contengan directrices sobre cómo detectar diferentes formas de desorden informativo.
- c) Modificar planes de estudio: Esta es una medida de fondo que apunta a hacer de la educación contra la desinformación un componente que se encuentra de manera transversal en los diferentes niveles educativos.

4.1.3. Prevención Cognitiva

Incluyen aquellas intervenciones proactivas que preparan a las personas contra la desinformación antes de que la encuentren. Este tipo de intervenciones se basan en la teoría de la inoculación cognitiva, según la cual, si se les advierte a los individuos que pueden estar siendo expuestos a alguna forma de desinformación y además se les expone a ejemplos atenuados en los que pueden ser desinformados, estos desarrollan una mayor inmunidad a la desinformación, o son menos propensos a propagarla.

4.1.4. Comunicación y Contranarrativa

Este grupo de estrategias agrupa diferentes técnicas comunicativas que tienen como objetivo hacer más persuasiva y comprensible la información que se ha comprobado veraz y pertinente.

- a) Adaptación a audiencias específicas: Esta estrategia consiste en moldear la información o los mensajes correctivos a las creencias o las identidades de las audiencias.
- b) Conocimiento experto: Este tipo de estrategias se centra en poner a los expertos, ya sean individuos o instituciones, al frente de la corrección de los mensajes desinformativos. También recoge actitudes no reactivas en las que los expertos divulgan información confiable a través de múltiples medios, como mensajes en plataformas, o la creación de contenido multimedia.
- c) Cuestionar/corregir fuentes: Desacreditar las fuentes de desinformación para tener un efecto más profundo y duradero, de tal modo que se evite que estas sigan actuando como proxis para obtener información confiable.
- d) Proporcionar explicaciones alternativas coherentes: Esta estrategia supone un paso adicional a la corrección de información, pues aduce la necesidad de proveer explicaciones relativamente robustas que expliquen por qué se rechaza la información, al tiempo que se ofrece una historia alternativa que reemplaza aquella que se ha desacreditado.
- e) Reducción de la exposición: En esta categoría se sitúan intervenciones que tienen como objetivo reducir la exposición de las audiencias a formas de desinformación. Las acciones concretas pueden involucrar reducir literalmente su visibilidad en las plataformas de redes sociales, o incluso remover la información por completo de las plataformas.
- f) Refutar afirmaciones completas: Consiste en hacer correcciones sobre la totalidad del mensaje considerado desinformativo, y no solo partes de este. Este ejercicio evita que las audiencias interpreten implícitamente que las partes no señaladas son de hecho correctas.
- g) Simplificar mensajes: Simplificar el léxico y formato de los mensajes correctivos. Así, se aconseja usar un mensaje claro y simple, sin añadir gráficos u otros elementos visuales.

4.1.5. Intervenciones Sistémicas

En este grupo se enmarcan medidas estructurales que modifican el ecosistema informativo a través de políticas, regulaciones y cambios en las plataformas digitales.

- a) Colaboración entre diferentes sectores: Consiste en robustecer las redes de trabajo entre los diferentes actores que están involucrados en la lucha contra la desinformación. Se destaca la colaboración entre las plataformas de difusión de información (medios tradicionales, medios nativos digitales, redes sociales, etc.), actores especializados en hacer verificación de información, instituciones educativas, e instituciones productoras de conocimiento (agencias estatales, think-tanks).

- b) Etiquetas: Esta estrategia supone marcar la información a manera de advertencia para que el usuario esté al tanto de que está frente a una potencial forma de desorden informativo. No hay consenso sobre cuál tipo de etiqueta es la más eficaz a la hora de combatir las formas de desorden informativo. Por un lado, algunos consideran que son preferibles las etiquetas específicas, las cuales señalan de manera inequívoca que la información es falsa (eg. Considerado Falso). Mientras tanto, otros aducen que son preferibles las etiquetas neutrales, que señalan sospechas sobre la información (eg. Contenido disputado).
- c) Regulación gubernamental: En esta categoría encontramos medidas que apuntan a crear regulaciones sobre el modo en que deben actuar las plataformas y otros actores en el ecosistema informativo. En particular, se proponen medidas para aumentar la transparencia en los algoritmos que determinan la información a la que están expuestos los usuarios, aumentar la moderación de contenido, o crear regulación que haga responsables a los actores maliciosos en el ecosistema de medios por sus acciones perjudiciales.

La ocurrencia de estas 19 estrategias en los artículos revisados se presenta en la siguiente tabla⁵:

Tabla 2. Artículos por tipo de estrategia analizada

Tipología	Estrategia	Referencias revisadas en las que se discute
Verificación y Análisis de Contenido	Accuracy nudging	Jalbert et al. 2025; Pennycook et al. 2020; Tu 2024.
	Fact-checking (automático)	Ahmed et al. 2022; Aïmeur, Amri, y Brassard 2023; Bondielli y Marcelloni 2019; Cantón-Correa et al. 2025; Conroy, Rubin, y Chen 2015; Ferreras Rodríguez y Sancho Belinchón 2025; Hakak et al. 2021; Kaliyar, Goswami, y Narang 2021; Lewandowsky, Ecker, y Cook 2017; Nasir, Khan, y Varlamis 2021; Sánchez Gonzales, González, y Alonso-González 2024; Ufarte Ruiz 2024; Zhang y Ghorbani 2020; Zhou y Zafarani 2019.
	Fact-checking (manual)	Aïmeur, Amri, y Brassard 2023; Bondielli y Marcelloni 2019; Cantón-Correa et al. 2025; Chung 2023; Lewandowsky y van der Linden 2021; Li, Li, y Yang 2025; Magallón-Rosa y Sánchez-Torres 2024; Reuter, Lee Hughes, y Buntain 2025, Sánchez Gonzales, González, y Alonso-González 2024; Sehl 2024; Zhang y Ghorbani 2020; Zhou y Zafarani 2019.
	Puntuaciones de credibilidad	Bondielli y Marcelloni 2019
	Estrategias implementadas por la comunidad	Jalbert et al. 2025

⁵ Es preciso notar que en algunas ocasiones una misma referencia aparece en más de una categoría. Esto se debe a que algunos de los artículos revisados analizan más de una estrategia para combatir el fenómeno desinformativo.

Educación y Alfabetización	Alfabetización digital	Ahmed et al. 2022; Carrillo y Montagut 2021; Chan 2024; Ferreras Rodríguez y Sancho Belinchón 2025; Irwanto et al. 2025; Jones-Jang y Kim 2018; Jones-Jang, Mortensen, y Liu 2019; Lewandowsky, Ecker, y Cook 2017; Moreno-Gil, Chaparro-Domínguez, y Pérez-Pereiro 2023; Magallón-Rosa y Sánchez-Torres 2024; McGowan-Kirsch y Quinlivan 2024; Moreno-Gil, Chaparro-Domínguez, y Pérez-Pereiro 2023; Reuter, Lee Hughes, y Buntain 2025; Sádaba y Salaverría 2023; Sehl 2024; Ufarte Ruiz 2024; Zhou y Zafarani 2019.
	Formación práctica	Moreno-Gil, Chaparro-Domínguez, y Pérez-Pereiro 2023; van der Linden, Roozenbeek, y Compton 2020.
	Guías	Tu 2024; Zhou y Zafarani 2019.
	Modificar planes de estudio	Magallón-Rosa y Sánchez-Torres 2024; Moreno-Gil, Chaparro-Domínguez, y Pérez-Pereiro 2023.
Prevención Cognitiva		Lewandowsky y van der Linden 2021; Lewandowsky et al. 2012; Lewandowsky, Ecker, y Cook 2017; McGowan-Kirsch y Quinlivan 2024; van der Linden, Roozenbeek, y Compton 2020.
Comunicación y Contranarrativa	Adaptación a audiencias específicas	Walter et al. 2020 ; Walter y Tukachinsky 2020.
	Conocimiento experto	Li, Li, y Yang 2025; Parviainen y Lahikainen 2019; Vraga y Bode 2017.
	Cuestionar/ corregir fuentes	Jalbert et al. 2025; Magallón-Rosa y Sánchez-Torres 2024; Walter y Tukachinsky 2020.
	Proporcionar explicaciones alternativas coherentes	Lewandowsky et al. 2012; McGowan-Kirsch y Quinlivan 2024; Walter y Tukachinsky 2020.
	Refutar afirmaciones completas	Jalbert et al. 2025; Vraga y Bode 2017; Walter et al. 2020.
	Simplificar mensajes	Lewandowsky et al. 2012; Walter et al. 2020.
	Reducción de la exposición	Reuter, Lee Hughes, y Buntain 2025
Intervenciones Sistémicas	Colaboración entre sectores	Moreno-Gil, Chaparro-Domínguez, y Pérez-Pereiro 2023; Walter et al. 2020.
	Etiquetas	Clayton et al. 2020; Zhang 2024
	Regulación gubernamental	Ahmed et al. 2022; Irwanto et al. 2025; Jones-Jang y Kim 2018; Lewandowsky, Ecker, y Cook 2017.

Fuente: Elaboración propia.

5. Discusión

A diferencia de las estrategias incluidas en el grupo denominado *Verificación y Análisis de Contenido*, las estrategias incluidas en las categorías de *Educación y Alfabetización* y *Prevención Cognitiva* se concentran en el desarrollo de habilidades proactivas que les permiten a las personas enfrentar la desinformación. La diferencia entre estos dos grupos radica en que el grupo 2 (*Educación y Alfabetización*) recoge esfuerzos de largo plazo orientados al desarrollo de habilidades de pensamiento crítico, en el que se ve a las personas como ciudadanos que deben interactuar con la información de manera autónoma. Por su parte, el grupo tres (*Prevención Cognitiva*) ofrece una caja de herramientas que las personas pueden adquirir y utilizar en su interacción con la información. La perspectiva, en este último caso, es más puntual que la del grupo 2.

En todo caso, lo que estos dos grupos de estrategias suponen es que el elemento crucial en la interacción con la información (veraz o no) son una serie de habilidades cognitivas que, una vez aprendidas, son implementadas en todas las interacciones con la información por igual. Algo semejante sucede con las estrategias del grupo 4 (*Comunicación y Contranarrativa*), pues no se discute si las audiencias son igualmente receptivas a las técnicas de comunicación y a la forma de presentar la información veraz, incluso cuando aquella está adaptada específicamente a las audiencias a las que se dirige.

Finalmente, la implementación de las estrategias del grupo 5 (*Intervenciones Sistémicas*) depende de todo un conjunto de factores políticos, económicos, sociales y culturales. Tales factores pueden determinar que la tendencia sea incluso contraria a lo que recomiendan los estudios que se enfocan en este tipo de estrategias. Por ejemplo, a comienzos del año 2025, Facebook e Instagram eliminaron los verificadores de información gestionados por organizaciones independientes, como ya lo había hecho la red social X. Las razones que están detrás de esta tendencia tienen connotaciones aparentemente políticas, pero también involucran la rentabilidad que ciertos flujos de información tienen para plataformas como Facebook, Instagram y X. Poner el énfasis en las estrategias incluidas en este grupo sin tener en cuenta la complejidad de estos factores políticos, económicos, etc., puede parecer poco realista y anodino.

Sin embargo (más allá de estos supuestos), buena parte de la literatura revisada identifica una serie de factores que inciden en la efectividad de las estrategias para combatir la desinformación. Por ejemplo, la posición política (Lewandowsky, Ecker y Cook 2017) o las creencias previas de las personas (Walter y Tukachinsky 2019), algunos rasgos de personalidad (Walter et al. 2020) estrategias heurísticas (Jones-Jang, Mortensen y Liu 2019), el grado y tipo de involucramiento afectivo con respecto a un tema (Carrillo y Montagut 2021; Zhang 2024). Estos factores parecen influir en la fuerza que puede tener determinada estrategia, así como en otras características de su efecto sobre la desinformación (por ejemplo, si este se da de manera diferenciada, si en algunos casos el efecto no se presenta, se atenúa o se incrementa). Este es el caso de la posición política previa de las personas, que produce variaciones en la efectividad de estrategias como el fact-checking cuando se analiza información consistente/inconsistente con dicha posición.

Se trataría, entonces, de factores moderadores. Sin embargo, parece que se trata de factores que no sólo aparecen incidentalmente en el uso de las estrategias analizadas, sino que se presentan de manera inevitable y cumplen un papel en nuestro uso de las estrategias y en nuestra interacción con la información. De ser así, probablemente se trataría de factores mediadores. Más aún, dado su papel decisivo en tales interacciones y el uso de las estrategias, puede que incluso se trate de factores determinantes.

Dos factores parecen ser cruciales al momento de considerar la efectividad de una estrategia para combatir la desinformación: se trata de los sesgos cognitivos (Jones-Jang, Mortensen y Liu 2019; Lewandowsky et al. 2012; Walter et al. 2020) y los afectos o emociones (Carrillo y Montagut 2021; Zhang 2024). En este orden de ideas resulta bastante peculiar que en la literatura revisada se insista en estrategias enfocadas en optimizar nuestras cogniciones o en cogniciones ya optimizadas. De este modo, un vacío en la literatura revisada es la escasa o nula consideración del papel que juegan factores como los sesgos cognitivos y las emociones en nuestra interacción con la información y, por lo tanto, en nuestra habilidad para utilizar estrategias que combatan la desinformación.

5.1. Verificación de información y la llegada de la inteligencia artificial

Una de las tendencias más relevantes observadas en la revisión de la literatura fue la importancia que tenía el fact-checking entre las estrategias para combatir las formas de desorden informativo. En particular, se encontró que más del 50% de los artículos revisados contemplaban las características, retos, posibilidades y medidas para hacer más efectiva esta estrategia. Entre las diferentes medidas contempladas, la más destacada era el desarrollo de métodos que permitieran automatizar los procesos de verificación (fact-checking automático). El problema de la automatización reviste tal relevancia entre los artículos revisados, que se hizo necesario diferenciar entre el chequeo manual y el automático a la hora de reportar los resultados de esta investigación. Ahora, el fact-checking, de manera general, se presenta como un método que, a pesar de que pueda ser usado para combatir las formas de desorden informativo, enfrenta obstáculos que limitan su efectividad. Los tres problemas más importantes que enfrenta el fact-checking es la eficacia del mensaje correctivo, la escalabilidad de la capacidad de corrección y el alcance que tienen las correcciones.

Uno de los retos más grandes que enfrenta el fact-checking es que, según múltiples estudios, las correcciones de las formas de desorden informativo no eliminan por completo su efecto (Walter et al. 2020; Walter y Tukachinsky 2020). Por un lado, la desinformación puede no lograr cambiar las creencias de las personas que fueron expuestas a la información falsa, más aún cuando estas correcciones entran en tensión con las creencias o la identidad de la audiencia. Por el otro, es posible que la desinformación siga afectando las actitudes y las acciones de sujetos de manera inconsciente. De este modo, se ha mostrado que la desinformación puede seguir afectando las actitudes y las acciones de las personas, incluso cuando estas crean en la veracidad de los mensajes correctivos (Lewandowsky et al. 2012).

Ante estas limitaciones, hay dos tipos de respuestas que no son excluyentes. Por un lado, hay un consenso entre los autores analizados de que la lucha contra la desinformación requiere de los esfuerzos y la coordinación de los diferentes actores involucrados en el ecosistema de medios, pues no existe una medida milagrosa contra la desinformación. En segundo lugar, están los análisis que se centran específicamente en hacer del fact-checking una estrategia más eficiente. Entre las medidas sugeridas se pueden encontrar acciones como: acompañar las correcciones de explicaciones coherentes que reemplacen el mensaje difuminado por el mensaje desinformador, evitar repetir la desinformación, conectar las correcciones con los sistemas de creencias de los usuarios, desacreditar las fuentes de desinformación y no solo el mensaje concreto, y corregir la desinformación de manera inmediata (Walter et al. 2020; Walter y Tukachinsky 2020).

Sin embargo, la última de estas recomendaciones, producir correcciones y hacerlas llegar a los usuarios de manera inmediata, supone un reto que ha sido puesto como uno de los más radicales a la hora de combatir las formas de desorden informativo usando fact-checking. El volumen de desinformación que circula en medios, particularmente en redes sociales es tal, que se hace casi imposible que los verificadores especializados puedan detectarla y mucho menos corregirla de manera inmediata. Este problema, según algunos, se ve exacerbado por la introducción de la IA generativa, que hace posible para los actores maliciosos crear desinformación mucho más sofisticada y difícil de detectar como falsa o engañosa.

Ante este panorama, actores interesados en hacer del fact-checking un recurso más efectivo han puesto de manifiesto la necesidad de adaptarse e introducir herramientas cada vez más sofisticadas en la detección y corrección de desinformación (Cantón-Correa et al. 2025; Sánchez Gonzales, González y Alonso-González 2024). Tal y como lo expresa Ana María Saavedra de Colombia Check, uno de los portales más importantes de corrección de información en Latinoamérica, “el fact checking y la alfabetización van en bici eléctrica y la desinformación en tren de alta velocidad” (Ferreras Rodríguez y Sancho Belinchón 2025).

Esta necesidad de introducir recursos más potentes que ayuden a hacer escalable el fact-checking se ve reflejada en los múltiples artículos de investigación que apuntaban hacia el desarrollo de mejores herramientas de verificación. En particular, es notable que la mayoría de los artículos de naturaleza experimental revisados se centren en el desarrollo de algoritmos que puedan detectar información errada (Hakak et al. 2021; Kaliyar, Goswami y Narang 2021; Nasir, Khan y Varlamis 2021). Así, el reto que se plantea en esta veta de investigación es de naturaleza técnica, pues el objetivo es crear algoritmos que puedan resolver un problema binario de clasificación, que consiste en clasificar noticias legítimas, como legítimas; y falsas, como falsas (Nasir, Khan y Varlamis 2021).

A pesar de que los artículos revisados reportan resultados positivos a la hora de desarrollar algoritmos más eficaces, aun es claro que hay obstáculos para la implementación del fact-checking automático. En primer lugar, las bases de datos para entrenar a los algoritmos no son lo suficientemente variadas ni amplias. En segundo, los métodos de detección automática pueden crear falsos positivos, dada la

dificultad que tienen para entender los matices del lenguaje natural. Tercero, dado que la automatización del fact-checking necesita del uso de IA, sus resultados pueden reflejar sesgos algorítmicos, de tal modo que puede conducir a resultados distorsionados, dañinos o injustos.

El tercer obstáculo que enfrentan los métodos de fact-checking automático se relaciona precisamente con su nivel de autonomía. A pesar de que los algoritmos desarrollados puedan ser más precisos al cumplir la tarea de clasificación, aún se requiere el acompañamiento de seres humanos. Así, en la literatura revisada se prioriza la aproximación denominada como *human-in-the-loop*. Esta necesidad se ve complementada por la manera en la que las organizaciones especializadas en la verificación de datos se aproximan a este tipo de soluciones. A pesar de que por ejemplo organizaciones especializadas en la verificación de datos usen la IA en sus actividades, aún existe recelo a la hora de automatizar los procesos (Sánchez Gonzales, González y Alonso-González 2024). Este panorama lleva a reconsiderar la escalabilidad de la verificación de datos. A pesar de los avances, parece haber un acuerdo en que seres humanos deben ser actores clave en la detección de información, lo cual, de manera inevitable, hace que el proceso de verificación no pueda enfrentar lo que muchos consideran es una crisis de la verdad y la información.

El cuarto obstáculo identificado se relaciona con el alcance de la corrección. Uno de los rasgos más preocupantes que tienen las formas de desorden informativo es el hecho de que se haga viral de manera relativamente fácil. Sin embargo, la información correctiva no suele disfruta de este carácter. En la presente revisión solo hubo un artículo que puso la atención en el problema del alcance en el centro de la discusión. En su trabajo, Myojung Chung (2024) estudia los factores que podrían aumentar las tasas de reposteo de las correcciones de mensajes desinformativos. En su trabajo, la autora concluye que es más probable que los usuarios reposten mensajes correctivos de información viral, por lo que los ejercicios de los verificadores deberían centrarse en este tipo de mensajes.

Este último problema es aún más acuciante en tanto que, en la actualidad, las formas de desorden informativo se están propagando por medio de aplicaciones de mensajería instantánea, como lo son WhatsApp, Facebook Messenger o Telegram (Newman et al. 2020: 19). Incluso, algunos ya ven con preocupación que chatbots basados en inteligencia artificial como Grok, ChatGPT, Gemini, o DeepSeek, los que poco a poco se han vuelto una fuente de noticias, especialmente entre la población de 18-44 años, se conviertan en actores que contribuyan a agudizar la crisis informativa en que nos encontramos (Orth y Bialik 2025). Aunque hasta ahora se considera que la desinformación es mucho más prevalente en las aplicaciones de mensajería instantánea, ellas comparten con este último grupo de fuentes de información una característica que podría dificultar el alcance de los verificadores, y es que estos difícilmente pueden acceder a los canales en los que la información es transmitida, por lo que verificar y corregir en estos medios se hace extremadamente difícil.

De este modo, se encuentra que la verificación de datos, a pesar de que sea la más popular entre las estrategias que encontramos, parece enfrentar un panorama no tan alentador, aun cuando se hayan dado avances notables en relación con su automatización. A pesar de que sea posible superar

problemas técnicos relacionados con la construcción de algoritmos que cumplan mejor con la tarea de clasificar mensajes según su autenticidad o falsedad, la efectividad de esta estrategia aún no está asegurada. Por un lado, el uso de algoritmos basados en IA no garantiza la aplicación de la estrategia de manera masiva a la mayoría de los mensajes relacionados con las formas de desorden informativo, pues aún se necesita de intervención humana (El problema de la escalabilidad), detectar desinformación no implica que podamos llegar a todas las personas que consumen (Problema del alcance), y finalmente, aunque la información correctiva pueda llegar a los usuarios que consumieron la desinformación, no es seguro que estos nuevos mensajes cambien sus creencias o que la desinformación no siga afectando sus actitudes o creencias (Problema de la eficacia).

6. Conclusiones

En los últimos años, el creciente impacto de la desinformación en la esfera pública ha sido considerado como el advenimiento de una crisis epistémica. En este contexto, el conocimiento experto se ha visto devaluado frente a la opinión, lo que debilita consensos sociales mínimos. Las diferentes formas de desorden informativo, ya sean producidas intencionalmente o no, están entrelazadas con la manipulación, la polarización y la erosión de la confianza en las instituciones. Todo ello constituye una amenaza cada vez mayor a la vida social y a los valores fundamentales de las culturas democráticas.

Este notable aumento de las formas de desorden informativo ha impulsado diversos esfuerzos de investigación académica y científica. Dichos esfuerzos analizan su naturaleza, los factores psicológicos y sociológicos que los caracterizan y, el que ha sido tema central de este trabajo, las estrategias que pueden utilizarse para combatirlos. En este trabajo se realizó un ejercicio de identificación y evaluación de algunos de los principales enfoques que profesionales de diversos campos han considerado medidas para limitar la información falsa, reducir las cámaras de eco y fortalecer la resiliencia cognitiva de los usuarios ante la desinformación.

Se pudo evidenciar que la verificación de datos es la estrategia más destacada para combatir la desinformación. Alrededor del 50 % de los artículos de mayor impacto en los últimos años se enfocan en analizar esta estrategia, en las distintas maneras en las que esta se puede expresar y en distintos aspectos de la misma. Sin embargo, resaltamos que se trata de una estrategia con limitaciones inherentes significativas. Uno de los principales retos para la verificación de datos es su eficacia limitada: los mensajes correctivos no siempre logran borrar por completo la influencia de la información engañosa, bien porque la velocidad con la que esta se difunde es mayor que la de estos, bien porque las características de forma y contenido de esta es más atractiva. De este modo, este tipo de información puede seguir afectando a las creencias y actitudes de forma subconsciente.

Entre los mayores desafíos para la verificación de datos se encuentra el que denominamos “problema de la escalabilidad”. Es tal el volumen de desorden informacional que la corrección sincrónica o inmediata por parte de verificadores humanos es prácticamente imposible. La descomunal

escala logística de este esfuerzo es una de las razones más fuertes para el uso de herramientas automatizadas de verificación de datos. Si bien estas herramientas ofrecen una solución, enfrentan sus propios obstáculos: en primer lugar, la insuficiencia de datos de entrenamiento, por los cuales estas herramientas no podrían lidiar con muchos tipos de información; en segundo lugar, el riesgo de falsos positivos o de falsos negativos, en los que información veraz podría ser considerada engañosa y viceversa; en tercer lugar, los sesgos algorítmicos resultantes de que estas herramientas usen IA.

Estos desafíos han llevado a pensar que la verificación totalmente automatizada no es fiable, por lo que la implementación de herramientas automatizadas de verificación de datos requiere inevitablemente supervisión humana. Sin embargo, esto produce dificultades adicionales, relacionadas con el entrenamiento especial que deben recibir quienes realicen esta supervisión, al tiempo que hace reaparecer dificultades conocidas, relacionadas con sesgos que podrían guiar la supervisión.

Una pregunta adicional que no se ha abordado en la literatura es cómo abordar la diferencia que existe entre la velocidad con la que se propaga la información engañosa y aquella con la que se difunden las correcciones. Podemos hablar de información engañosa que se viraliza, pero no es usual encontrar correcciones viralizadas. Se trata de una diferencia esencial para el alcance de la verificación y corrección de información.

Lo anterior no deja duda de que la verificación de datos no puede ser la única estrategia para combatir el desorden informacional (incluso la principal estrategia). Este hecho lleva a que sobresalgan otras estrategias, orientadas, por ejemplo, al desarrollo de habilidades cognitivas que permitan a las personas tener una interacción proactiva (más que reactiva) con la información. Algunas de estas estrategias implican objetivos más extendidos en el tiempo, como el desarrollo del pensamiento crítico, mientras que otras buscan objetivos alcanzables en plazos más cortos, como la adquisición de conjuntos de herramientas específicas para la interacción con la información.

También es importante subrayar que la literatura sobre estrategias para combatir el desorden informacional reconoce una serie de factores que invariablemente influyen en la eficacia de cualquiera de las estrategias analizadas. Esos factores incluyen sobre todo creencias preexistentes de diferentes órdenes (políticas, científicas, religiosas, éticas, etc.), rasgos de personalidad (por ejemplo, extraversión o introversión, empatía, adaptabilidad), heurísticas (como disponibilidad, representatividad, anclaje) y, en particular, su implicación afectiva o emocional con un tema. Queda por ver la manera en la que estos factores inciden en la eficacia de las estrategias en cuestión, es decir, si se trata de factores moderadores, mediadores o determinantes. Esto último es un asunto que no se ha considerado hasta ahora en la literatura mencionada.

Como resultado de este importante vacío, se da una escasa consideración a elementos como sesgos cognitivos y emociones, pese a que se les reconoce como factores incidentes en la eficacia de las estrategias. La consecuencia de esto es que muchas de las estrategias analizadas se centren en optimizar procesos cognitivos (como en el caso del fortalecimiento del pensamiento crítico y la

alfabetización informacional, digital y mediática) o mejorar la confiabilidad de la información con la que interactuamos (como en el caso del énfasis en la verificación de información, ya sea manual o automática).

Las limitaciones de nuestra revisión tienen que ver principalmente con su alcance, pues no podemos afirmar que se trate de una revisión exhaustiva de la literatura existente. Además, el campo de investigación es tan dinámico que seguramente hayan aparecido más estudios relevantes entre el momento en el que se hizo la revisión (junio-agosto de 2025) y el momento de su publicación. Sin embargo, confiamos en que los nuevos aportes a la discusión no cambien drásticamente las tendencias dentro de esta.

Las investigaciones futuras deberían aclarar la escalabilidad y alcance posibles para la verificación automática de hechos. Del mismo modo, se requiere más investigación sobre cómo integrar en las aproximaciones los sesgos emocionales y cognitivos, así como otros factores que siguen sin explorarse suficientemente, buscando determinar si se trata de factores mediadores, moderadores o determinantes de la eficiencia de cada estrategia o grupo de estrategias. Esto permitirá una comprensión mucho más robusta de lo que debe y puede hacerse para combatir las formas de desorden informacional.

Bibliografía

- Ahmed, Mohiuddin, Sascha-Dominik Bachmann, Chris Martin, Tristram Walker, James Rooyen, y Abu Barkat. 2022. "False Information as a Threat to Modern Society: A Systematic Review of False Information, Its Impact on Society, and Current Remedies". *Journal of Information Warfare* 21 (2): 105–20. <https://www.jstor.org/stable/27199973>
- Aïmeur, Esma, Sabrine Amri, y Gilles Brassard. 2023. "Fake News, Disinformation and Misinformation in Social Media: A Review". *Social Network Analysis and Mining* 13 (1): 1–36. <https://doi.org/10.1007/s13278-023-01028-5>
- Atlas.ti. 2025. "What is a Critical Literature Review?" Atlas.ti, 2025. <https://atlasti.com/guides/literature-review/critical-literature-review#:~:text=A%20critical%20literature%20review%20is,implicit%20assumptions%20in%20current%20research>.
- Bondielli, Alessandro, y Francesco Marcelloni. 2019. "A Survey on Fake News and Rumour Detection Techniques". *Information Sciences* 497: 38–55. <https://doi.org/10.1016/j.ins.2019.05.035>
- Cantón-Correa, Javier, Andrés Montoro-Montaroso, Juan Gómez-Romero, y Miguel Molina-Solana. 2025. "Perfiles y necesidades formativas de los fact-checkers de la Península Ibérica: impacto de la IA." *Doxa Comunicación. Revista Interdisciplinar De Estudios De Comunicación Y Ciencias Sociales* 40: 465–91. <https://doi.org/10.31921/doxacom.n40a2725>
- Carrillo, Nereida, y Marta Montagut. 2021. "Tackling online disinformation through media literacy in Spain: The project 'Que no te la cuelen'". En *Catalan Journal of Communication & Cultural Studies*, 13: 149–57. https://doi.org/10.1386/cjcs_00044_7
- Chan, Michael. 2024. "News literacy, fake news recognition, and authentication behaviors after exposure to fake news on social media". *New Media & Society* 26 (8): 4669–88. <https://doi.org/10.1177/14614448221127675>
- Chung, Myojung. 2024. "Share to Stop the Harm: How Social Media Metrics Drive Sharing of Fact-Checking Messages via First-Person Perception". *Mass Communication and Society* 27 (5): 925–48. <https://doi.org/10.1080/15205436.2023.2240302>
- Clayton, Katherine, Spencer Blair, Jonathan A. Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, et al. 2020. "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media". *Political Behavior* 42 (4): 1073–95. <https://doi.org/10.1007/s11109-019-09533-0>
- Conroy, Niall J., Victoria L. Rubin, y Yimin Chen. 2015. "Automatic deception detection: Methods for finding fake news". *Proceedings of the ASIST Annual Meeting* 52 (1): 1–4. <https://doi.org/10.1002/pra2.2015.145052010082>
- Ferreras Rodríguez, Eva M., y Celia Sancho Belinchón. 2025. "Propuestas de alfabetización mediática desde el fact checking profesional en Latinoamérica". *Revista de Comunicación* 24 (1): 205–28. <https://doi.org/10.26441/RC24.1-2025-3777>

- Grant, Maria J., y Andrew Booth. 2009. "A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies." *Health Information and Libraries Journal* 26 (2): 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Hakak, Saqib, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, y Wazir Zada Khan. 2021. "An ensemble machine learning approach through effective feature extraction to classify fake news". *Future Generation Computer Systems* 117 (2021): 47–58. <https://doi.org/10.1016/j.future.2020.11.022>
- Herrero, Esperanza, y Susana Herrera Damas. 2021. "El fact-checker en español alrededor del mundo: Perfil, similitudes y diferencias entre verificadores hispanohablantes". *Revista de Comunicación de la SEECI* 54: 49–77. <https://doi.org/10.15198/seeci.2021.54.e725>
- Irwanto, Irwanto, Tuti Bahfiarti, Andi Alimuddin Unde, y Alem Febri Sonni. 2025. "Information disorder's impact on adolescents: publication trends and recommendations". *Frontiers in Communication* 10: 1–16. <https://doi.org/10.3389/fcomm.2025.1495536>
- Jalbert, Madeline, Morgan Wack, Pragya Arya, y Luke Williams. 2025. "Social truth queries: Development of a new user-driven intervention for countering online misinformation". *Journal of Applied Research in Memory and Cognition* 14 (2): 205–17. <https://doi.org/10.1037/mac0000142>
- Jones-Jang, S M, T Mortensen, y J Liu. 2019. "Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't". *American Behavioral Scientist* 65 (2): 371–88. <https://doi.org/10.1177/0002764219869406>
- Jones-Jang, S. Mo, y Joon K. Kim. 2018. "Third Person Effects of Fake News: Fake News Regulation and Media Literacy Interventions". *Computers in Human Behavior* 80 (2018): 295–302. <https://doi.org/10.1016/j.chb.2017.11.034>
- Kaliyar, Rohit Kumar, Anurag Goswami, y Pratik Narang. 2021. "FakeBERT: Fake News Detection in Social Media with a BERT-Based Deep Learning Approach". *Multimedia Tools and Applications* 80 (8): 11765–88. <https://doi.org/10.1007/s11042-020-10183-2>
- Lewandowsky, Stephan, Ullrich K. H. Ecker, y John Cook. 2017. "Beyond Misinformation: Understanding and Coping with the 'Post-Truth' Era". *Journal of applied research in memory and cognition* 6 (4): 353–69. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- Lewandowsky, Stephan, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, y John Cook. 2012. "Misinformation and Its Correction: Continued Influence and Successful Debiasing". *Psychological Science in the Public Interest* 13 (3): 106–31. <https://doi.org/10.1177/1529100612451018>
- Lewandowsky, Stephan, y Sander van der Linden. 2021. "Countering Misinformation and Fake News Through Inoculation and Prebunking". *European Review of Social Psychology* 32 (2): 348–84. <https://doi.org/10.1080/10463283.2021.1876983>
- Li, Gaofei, Mengyu Li, y Sijia Yang. 2025. "The 'Whole-Of-Society' Approach for Misinformation Correction: How Expert Didactic TikTok Videos Motivate Citizen Fact-Checking and Vaccine Promotion". *Journal of Health Communication* 30 (10–12): 356–68. <https://doi.org/10.1080/10810730.2025.2503179>
- Magallón-Rosa, Raúl, y Mabel Sánchez-Torres. 2024. "Disinformation and climate crisis: Challenges and proposals for the future". *Catalan Journal of Communication and Cultural Studies* 16 (2): 167–76. https://doi.org/10.1386/cjcs_00107_7
- McGowan-Kirsch, Angela M., y Grace V. Quinlivan. 2024. "Educating emerging citizens: Media literacy as a tool for combating the spread of image-based misinformation". *Communication Teacher* 38 (1): 41–52. <https://doi.org/10.1080/17404622.2023.2271548>
- McIntyre, Lee. 2023. *On Disinformation: How to Fight for Truth and Protect Democracy*. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/14889.001.0001>
- Moreno-Gil, Victoria, María-Ángeles Chaparro-Domínguez, y Marta Pérez-Pereiro. 2023. "Future journalists' fight against disinformation: analysis of university training offers and challenges in the Spanish context". *Communication & Society* 36 (2): 171–85. <https://doi.org/10.15581/003.36.2.171-185>
- Moreno-Gil, Victoria, Xavier Ramon, y Ruth Rodríguez-Martínez. 2021. "Fact-Checking Interventions as Counteroffensives to Disinformation Growth: Standards, Values, and Practices in Latin America and Spain". *Media and Communication* 9 (1): 251–63. <https://doi.org/10.17645/mac.v9i1.3443>
- Nasir, Jamal Abdul, Osama Subhani Khan, y Iraklis Varlamis. 2021. "Fake news detection: A hybrid CNN-RNN based deep learning approach". *International journal of information management data insights* 1 (1): 1–12. <https://doi.org/10.1016/j.jiimei.2020.100007>
- Navarro-Sierra, Nuria, Silvia Magro-Vela, y Raquel Vinader-Segura. 2024. "Research on Disinformation in Academic Studies: Perspectives through a Bibliometric Analysis". *Publications* 12 (2). <https://doi.org/10.3390/publications12020014>
- Newman, Nic, Richard Fletcher, Anne Schulz, Simge Andi, y Rasmus Kleis Nielsen. 2020. *Reuters Institute Digital News Report 2020*. Reuters Institute for the Study of Journalism. <https://doi.org/10.60625/RISJ-048N-AP07>
- Orth, Taylor, y Carl Bialik. 2025. *Trust in Media 2025: Which news sources Americans use and trust*. YouGov. <https://today.yougov.com/politics/articles/52272-trust-in-media-2025-which-news-sources-americans-use-and-trust>
- Parviainen, Jaana, y Lauri Lahikainen. 2019. "Negative expertise in conditions of manufactured ignorance: epistemic strategies, virtues and skills". *Synthese* 198 (4): 3873–91. <https://doi.org/10.1007/s11229-019-02315-5>
- Pennycook, Gordon, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, y David G. Rand. 2020. "Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention." *Psychological Science* 31 (7): 770–80. <https://doi.org/10.1177/0956797620939054>

- Reuter, Christian, Amanda Lee Hughes, y Cody Buntain. 2025. "Combating information warfare: state and trends in user-centred countermeasures against fake news and misinformation". *Behaviour & Information Technology* 44 (13): 3348–61. <https://doi.org/10.1080/0144929X.2024.2442486>
- Sádaba, Charo, y Ramón Salaverría. 2023. "Combatar 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>
- Sánchez Gonzales, Hada M, María Sánchez González, y Marián Alonso-González. 2024. "Artificial intelligence and disinformation literacy programmes used by European fact-checkers". *Catalan Journal of Communication & Cultural Studies* 16 (2): 237–55. https://doi.org/10.1386/cjcs_00111_1
- Schissler, Matt. 2025. "Beyond Hate Speech and Misinformation: Facebook and the Rohingya Genocide in Myanmar". *Journal of Genocide Research* 27 (3): 445–70. <https://doi.org/10.1080/14623528.2024.2375122>
- Sehl, Annika. 2024. "Public service media as pivotal in combating misinformation and disinformation: prerequisites and approaches". *European journal of communication* 39 (6): 582–94. <https://doi.org/10.1177/02673231241294185>
- Tenove, Chris. 2020. "Protecting Democracy from Disinformation: Normative Threats and Policy Responses". *The International Journal of Press/Politics* 25 (3): 517–37. <https://doi.org/10.1177/1940161220918740>
- Tu, Fangjing. 2024. "Empowering social media users: nudge toward self-engaged verification for improved truth and sharing discernment". *Journal of Communication* 74 (3): 225–36. <https://doi.org/10.1093/joc/jqae007>
- Ufarte Ruiz, María José. 2024. "The consolidation of disinformation as a discipline: Reflections, challenges and perspectives for future action". *Catalan Journal of Communication & Cultural Studies* 16 (2): 159–66. https://doi.org/10.1386/cjcs_00106_7
- van der Linden, Sander, Jon Roozenbeek, y Josh Compton. 2020. "Inoculating Against Fake News About COVID-19". *Frontiers in Psychology* 11 (2020): 1–7.
- Vraga, Emily K., y Leticia Bode. 2017. "Using Expert Sources to Correct Health Misinformation in Social Media". *Science Communication* 39 (5): 621–45. <https://doi.org/10.1177/1075547017731776>
- Walter, Nathan, Jonathan Cohen, R. Lance Holbert, y Yasmin Morag. 2020. "Fact-Checking: A Meta-Analysis of What Works and for Whom". *Political Communication* 37 (3): 350–75. <https://doi.org/10.1080/10584609.2019.1668894>
- Walter, Nathan, y Riva Tukachinsky. 2020. "A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It?" *Communication Research* 47 (2): 155–77. <https://doi.org/10.1177/0093650219854600>
- Wardle, Claire, y Hossein Derakhshan. septiembre de 2017. *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe report DGI. <https://rm.coe.int/information-disorderreportnovember-2017/1680764666>
- World Health Organization. 2020. "Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. Joint statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC". World Health Organization. <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>
- Zhang, Bingbing. 2024. "Examining the effectiveness of social media warning labels: The role of worldview inconsistency and psychological reactanc". *The Journal of Social Media in Society* 13 (2): 131–50.
- Zhang, Xichen, y Ali A. Ghorbani. 2020. "An overview of online fake news: Characterization, detection, and discussion". *Information Processing & Management* 57 (2): 1–26. <https://doi.org/10.1016/j.ipm.2019.03.004>
- Zhou, Xinyi, y Reza Zafarani. 2019. "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities". *ACM Comput. Surv. (New York, NY, USA)* 53 (5). <https://doi.org/10.1145/3395046>

