J-40402082-9

Fundación Aula Virtual

Aula Virtual

Generando Conocimiento

http://www.aulavirtual.web.ve

ISSN: 2665-0398

Deposito Legal: LA2020000026

Vol. 7 Nº 14 Año 2026

Periodicidad Continua

ISSN: 2665-0398

AulaVirtual
Generando Conocimiento

REVISTA CIENTÍFICA AULA VIRTUAL

Revista Científica Arbitrada de Fundación Aula Virtual

Email: revista@aulavirtual.web.ve
URL: http://aulavirtual.web.ve/revista

ISSN: 2665-0398
Depósito Legal: LA2020000026
País: Venezuela
Año de Inicio: 2020
Periodicidad: Continua
Sistema de Arbitraje: Revisión por pares. "Doble Ciego"
Licencia: **Creative Commons** CC BY NC ND
Volumen: 7
Número: 14
Año: 2026
Período: Enero 2026 - Junio 2026 (continua)
Dirección Fiscal: Av. Libertador, Arca del Norte, Nro. 52D, Barquisimeto estado Lara, Venezuela, C.P. 3001

La Revista seriada Científica Arbitrada e Indexada **Aula Virtual**, es de acceso abierto y en formato electrónico; la misma está orientada a la divulgación de las producciones científicas creadas por investigadores en diversas áreas del conocimiento. Su cobertura temática abarca Tecnología, Ciencias de la Salud, Ciencias Administrativas, Ciencias Sociales, Ciencias Jurídicas y Políticas, Ciencias Exactas y otras áreas afines. Su publicación es **CONTINUA**, indexada y arbitrada por especialistas en el área, bajo la modalidad de doble ciego. Se reciben las producciones tipo: *Artículo Científico* en las diferentes modalidades cualitativas y cuantitativas, *Avances Investigativos*, *Ensayos, Reseñas Bibliográficas, Ponencias o publicaciones derivada de eventos*, y cualquier otro tipo de investigación orientada al tratamiento y profundización de la información de los campos de estudios de las diferentes ciencias. La Revista **Aula Virtual,** busca fomentar la divulgación del conocimiento científico y el pensamiento crítico reflexivo en el ámbito investigativo.

AulaVirtual
Generando Conocimiento

ISSN: 2665-0398

DESINFORMACIÓN EN ENTORNOS CIFRADOS Y PREVENCIÓN DE CONFLICTOS SOCIALES: UNA REVISIÓN SISTEMÁTICA DE ESTRATEGIAS MULTISECTORIALES

DISINFORMATION IN ENCRYPTED ENVIRONMENTS AND SOCIAL CONFLICT PREVENTION: A SYSTEMATIC REVIEW OF MULTISECTORAL STRATEGIES

Tipo de Publicación: Articulo Científico
Recibido: 04/06/2026
Aceptado: 05/07/2026
Publicado: 30/06/2026
Código Único AV: e779
Páginas: 1(1752-1775)
DOI: https://doi.org/10.5281/zenodo.21209952

Autores:
Himbler Jacyson Aceval Cienfuegos
Comunicador Social
Maestro en Comunicación para el Desarrollo
https://orcid.org/0000-0002-5437-7765
E-mail: haceval@unheval.edu.pe
Afiliación: Universidad Nacional Hermilio Valdizán
País: República del Perú

Eddie Misael Samaniego Pimentel
Comunicador Social
Magíster en Gestión Pública
https://orcid.org/0009-0005-7891-961X
E-mail: esamaniego@unheval.edu.pe
Afiliación: Universidad Nacional Hermilio Valdizán
País: República del Perú

Ivan Agui Guillen
Licenciado en Comunicación Social
Maestro en Educación, mención Investigación y Docencia Superior
https://orcid.org/0000-0002-0730-2234
E-mail: ivan.agui@udh.edu.pe
Afiliación: Universidad de Huánuco
País: República del Perú

Rocío Verónica Rasmuzzen Santamaria
Licenciada en Administración
Doctora en Administración
https://orcid.org/0000-0001-8772-9360
E-mail: rrasmuzzen@unheval.edu.pe
Afiliación: Universidad Nacional Hermilio Valdizán
País: República del Perú

Resumen

La desinformación en plataformas de mensajería privada constituye un desafío crítico para las sociedades contemporáneas, debido a que los entornos cifrados facilitan la circulación opaca de contenidos falsos, rumores y narrativas polarizantes con potencial para erosionar la confianza pública y escalar conflictos sociales. La relevancia de este problema se intensifica en contextos electorales, sanitarios, bélicos y de crisis institucional, donde WhatsApp, Telegram, Line y otros servicios de mensajería funcionan tanto como canales de comunicación interpersonal como espacios de movilización colectiva. El objetivo de este artículo fue evaluar las estrategias implementadas por plataformas de mensajería privada, gobiernos y sociedad civil para contrarrestar la desinformación en entornos cifrados y examinar su efectividad en la prevención de conflictos sociales entre 2015 y 2026. Se desarrolló una revisión sistemática de literatura científica, siguiendo criterios de búsqueda, inclusión y exclusión orientados a seleccionar estudios sobre intervenciones técnicas, regulatorias y comunitarias frente a la desinformación en aplicaciones de mensajería privada. Los resultados evidencian que las plataformas han priorizado bots de verificación, chatbots, sistemas de advertencia, huellas digitales de contenido y moderación descentralizada; los gobiernos han recurrido a marcos regulatorios, contranarrativas, transparencia y comunicación oficial; mientras que la sociedad civil ha impulsado fact-checking, alfabetización mediática, tiplines de verificación, campañas comunitarias y estrategias híbridas de comunicación. Se concluye que la efectividad de estas estrategias es diferenciada e indirecta, pues depende de la confianza del usuario, la legitimidad institucional, la adaptación cultural y la compatibilidad con la privacidad. La prevención de conflictos sociales exige, por tanto, modelos colaborativos que integren tecnología, regulación proporcional y resiliencia comunitaria.

Palabras Clave Desinformación digital, mensajería privada, entornos cifrados, verificación de hechos, conflictos sociales.

Abstract

Disinformation on private messaging platforms represents a critical challenge for contemporary societies, as encrypted environments enable the opaque circulation of false content, rumors, and polarizing narratives with the potential to undermine public trust and escalate social conflicts. This problem becomes particularly relevant in electoral, health, war-related, and institutional crisis contexts, where WhatsApp, Telegram, Line, and other messaging services operate both as interpersonal communication channels and as spaces for collective mobilization. The objective of this article was to evaluate the strategies implemented by private messaging platforms, governments, and civil society to counter disinformation in encrypted environments and to examine their effectiveness in preventing social conflicts between 2015 and 2026. A systematic literature review was conducted, following structured search, inclusion, and exclusion criteria aimed at identifying studies on technical, regulatory, and community-based interventions against disinformation in private messaging applications. The findings show that platforms have prioritized verification bots, fact-checking chatbots, warning systems, content fingerprinting, and decentralized moderation; governments have adopted regulatory frameworks, counternarratives, transparency measures, and official communication strategies; while civil society has promoted fact-checking, media literacy, verification tiplines, community campaigns, and hybrid communication strategies. The review concludes that the effectiveness of these strategies is differentiated and indirect, as it depends on user trust, institutional legitimacy, cultural adaptation, and compatibility with privacy protections. Therefore, the prevention of social conflicts requires collaborative models that integrate technological innovation, proportional regulation, and community resilience. This study contributes to the field by offering an integrated understanding of how different actors respond to disinformation in encrypted environments and by highlighting the need for context-sensitive interventions that protect digital rights while reducing the social harms associated with false information.

Keywords Digital disinformation, private messaging, encrypted environments, fact-checking, social conflicts.

Introducción

La desinformación en entornos digitales constituye uno de los fenómenos más intrincados y desestabilizadores de las sociedades contemporáneas, cuya incidencia se ha agudizado con la expansión de plataformas de mensajería privada como WhatsApp y Telegram. Estos servicios, caracterizados por el cifrado de extremo a extremo y por la limitada intervención de mecanismos centralizados de moderación, han configurado ecosistemas comunicativos opacos en los que los contenidos falsos circulan con escasos dispositivos de contención. Esta dinámica produce efectos relevantes en la polarización social, la erosión de la confianza pública y el posible escalamiento de conflictos colectivos (Chagas & Da-Costa, 2023). En este escenario, la evaluación de las estrategias desplegadas por plataformas tecnológicas, gobiernos y organizaciones de la sociedad civil para contrarrestar la desinformación en entornos cifrados adquiere especial pertinencia, sobre todo cuando se analiza su efectividad en la prevención de conflictos sociales.

El estudio de la desinformación en plataformas de mensajería privada se sostiene en marcos teóricos que articulan las dinámicas comunicativas propias de los entornos cifrados con los procesos de formación de opinión pública, confianza interpersonal y conflictividad social. Duarte & Rosa (2023) plantean un análisis del ecosistema de la desinformación desde una perspectiva multidimensional, al señalar que las redes sociales y los servicios de mensajería instantánea, como WhatsApp y Telegram, operan como amplificadores de contenidos falsos. En estos espacios, la veracidad atribuida entre pares, la credibilidad derivada de los vínculos de confianza y la emocionalidad adquieren un papel decisivo en el consumo informativo y en la viralización de mensajes. Esta aproximación resulta cardinal para comprender por qué las estrategias de contención de la desinformación enfrentan dificultades específicas en entornos privados, a diferencia de las plataformas abiertas.

De manera complementaria, Gursky et al., (2022) desarrollan el concepto de lógica de cascada (*cascade logic*) para explicar los mecanismos mediante los cuales la información transita entre espacios privados y públicos en las aplicaciones de mensajería. Desde esta perspectiva, las aplicaciones de chat operan en una zona liminar entre la publicidad convencional de las redes sociales y la privacidad de la comunicación interpersonal. Por ello, constituyen espacios donde la desinformación no solo se disemina, sino que también puede planificarse, coordinarse y redistribuirse hacia plataformas abiertas. Este marco teórico permite advertir la naturaleza multiplataforma de las campañas de desinformación y, en consecuencia, la necesidad de diseñar estrategias de contención igualmente transversales.

ISSN: 2665-0398

Por su parte, Díez-Garrido et al., (2021) aportan una perspectiva centrada en las affordances técnicas de las plataformas de mensajería, las cuales dificultan la verificación de contenidos. Los autores sostienen que la privacidad de los espacios de comunicación creados en estas plataformas constituye uno de los principales obstáculos para la detección y contención de información falsa. Asimismo, evidencian que estos servicios han sido instrumentalizados en procesos electorales de diversos países, lo que refuerza la necesidad de promover mecanismos de detección de noticias falsas y prácticas de fact-checking orientadas a preservar una percepción pública de la realidad basada en información verificable.

La literatura reciente ha documentado avances sustantivos en la comprensión de las dinámicas de desinformación en plataformas de mensajería privada, así como en el análisis de las respuestas institucionales frente a este fenómeno. Melo et al., (2024) realizaron un estudio de aproximadamente diez millones de mensajes procedentes de 1.101 grupos públicos de WhatsApp dedicados a la discusión política en Brasil, con el propósito de evaluar la efectividad de las medidas implementadas por la plataforma para restringir la diseminación masiva de contenidos. Sus hallazgos revelan que dichas medidas pueden ser eludidas con relativa facilidad, pues el 59 % del contenido duplicado identificado como reenviado muchas veces no recibió la etiqueta correspondiente. Este estudio aporta evidencia empírica relevante sobre las limitaciones de las estrategias técnicas adoptadas por las plataformas digitales.

En una línea afín, Krishnan et al., (2021) examinaron las respuestas de doce plataformas de redes sociales y mensajería ante la desinformación relacionada con la COVID-19. Los autores concluyeron que, si bien plataformas como Facebook, Instagram, YouTube y Twitter presentaron respuestas mayormente consistentes, otras plataformas, incluidas algunas de mensajería, mostraron variaciones significativas respecto de los tipos de contenido prohibido, los criterios utilizados para intervenir y los mecanismos diseñados para mitigar la desinformación. Este hallazgo evidencia la fragmentación de las respuestas existentes y plantea la necesidad de establecer estándares generales entre plataformas, a fin de abordar la desinformación de forma más cohesionada.

Asimismo, Pasquetto et al., (2022) exploraron el potencial de WhatsApp como herramienta para la corrección de información falsa. Su investigación muestra que las affordances específicas de la plataforma, entre ellas la flexibilidad de formatos y la posibilidad de seleccionar audiencias concretas, pueden combinarse con el capital social de los usuarios para incrementar la redistribución de mensajes de verificación. Los resultados indican que los usuarios tienden a compartir desmentidos

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

con mayor frecuencia cuando los reciben de personas cercanas o de individuos con afinidades políticas semejantes. Esta evidencia abre una perspectiva sugerente, pues sitúa a las plataformas de mensajería no solo como vectores de desinformación, sino también como espacios potencialmente útiles para su contención.

A pesar de los avances descritos, la literatura presenta vacíos significativos que justifican el desarrollo de una revisión sistemática. Chagas & Da-Costa (2023) identifican la opacidad ambiental como una condición estructural de los servicios de mensajería instantánea basados en sistemas de cifrado de extremo a extremo. Según los autores, dicha condición incide en la ética y la transparencia de la investigación académica centrada en WhatsApp y otros servicios de mensajería. Además, sostienen que países emergentes con grandes bases de usuarios, como Brasil e India, han experimentado efectos negativos asociados con la adopción de WhatsApp por grupos políticamente orientados. Sin embargo, también reconocen la persistente insuficiencia de medidas orientadas a fortalecer la transparencia de las plataformas y facilitar la investigación científica.

Por otra parte, Martínez et al., (2022) advierten que el estudio de la incidencia de las plataformas de mensajería en la comunicación política permanece escasamente explorado en comparación con la amplia producción académica sobre redes sociales abiertas. Su investigación sobre WhatsApp como "caja negra" revela la posible presencia de filtros burbuja importados desde otros espacios digitales, así como la centralidad de emociones negativas en las interacciones políticas. No obstante, los autores reconocen que la comprensión de estas dinámicas exige un desarrollo teórico y metodológico todavía incipiente.

Finalmente, Cocha et al., (2024) subrayan que el abordaje de la desinformación en la era digital, especialmente durante conflictos globales, requiere un enfoque multifacético que involucre a plataformas tecnológicas, gobiernos, sociedad civil y sector académico. Sin embargo, también reconocen que la comprensión de las dinámicas específicas de la desinformación en contextos particulares continúa siendo limitada. Ello evidencia la necesidad de estudios que sinteticen, desde una perspectiva comparativa, las estrategias desplegadas por los distintos actores involucrados en la lucha contra la desinformación en entornos cifrados.

A partir de los vacíos identificados, el presente artículo tiene como objetivo evaluar las estrategias implementadas por plataformas de mensajería privada, gobiernos y organizaciones de la sociedad civil para contrarrestar la desinformación en entornos cifrados, así como examinar su efectividad en la prevención de conflictos sociales durante el período 2015-2026.

ISSN: 2665-0398

Metodología

La presente investigación se desarrolló bajo el enfoque de una revisión sistemática de la literatura científica, siguiendo las directrices establecidas por el protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

La búsqueda bibliográfica se realizó exclusivamente en la base de datos Scopus. La selección de esta única base de datos obedece a tres razones fundamentales. En primer lugar, Scopus constituye la mayor base de datos de literatura revisada por pares a nivel mundial. En segundo lugar, ofrece herramientas avanzadas de búsqueda booleana y filtrado que permiten delimitar con precisión los resultados. En tercer lugar, la indexación rigurosa de Scopus garantiza un estándar de calidad en los artículos recuperados.

La estrategia de búsqueda se operacionalizó mediante la siguiente fórmula booleana, diseñada para capturar la intersección temática entre desinformación, plataformas de mensajería privada, estrategias de contención y conflictos sociales: (("disinformation" OR "misinformation" OR "fake news" OR "false information") AND ("messaging app*" OR "WhatsApp" OR "Telegram" OR "Signal" OR "encrypted messaging" OR "end-to-end encryption" OR "private messaging") AND ("counter*" OR "combat*" OR "strateg*" OR "intervention*" OR "regulation*" OR "fact-check*" OR "content moderation" OR "media literacy"))

Para guiar la revisión sistemática y estructurar el análisis de los estudios seleccionados, se formularon las siguientes preguntas de investigación: PI1: ¿Qué estrategias técnicas han implementado las plataformas de mensajería privada (WhatsApp, Telegram, Signal) para contrarrestar la propagación de desinformación en entornos cifrados, y qué evidencia existe sobre su efectividad? PI2: ¿Qué marcos regulatorios y políticas gubernamentales se han desarrollado para abordar la desinformación en plataformas de mensajería cifrada, y cómo equilibran la tensión entre seguridad pública y privacidad de las comunicaciones? PI3: ¿Qué iniciativas de la sociedad civil —incluyendo organizaciones de verificación de hechos, medios de comunicación y organizaciones no gubernamentales— han demostrado ser efectivas para mitigar la desinformación en entornos de mensajería privada y prevenir el escalamiento de conflictos sociales?

Criterio de inclusión	Criterio de exclusión
Estudios sobre desinformación en mensajería privada o entornos cifrados.	Estudios centrados solo en plataformas abiertas sin referencia a mensajería privada.
Artículos originales o de revisión publicados en revistas con revisión por pares.	Editoriales, cartas, resúmenes de conferencia, capítulos de libro o literatura gris.
Estudios sobre intervenciones de Plataformas, gobiernos o sociedad civil.	Estudios exclusivamente técnico-computacionales sin Enfoque social, político o de conflicto.
Estudios con texto completo disponible.	Estudios sin acceso a texto completo.

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

Criterio de inclusión	Criterio de exclusión
Registros identificados en bases de datos o mediante revisión manual de referencias.	Artículos duplicados durante el cribado.

Tabla 1. Criterios de elegibilidad

La aplicación secuencial de estos criterios de inclusión y exclusión permitió conformar un corpus de estudios pertinentes y de calidad para responder al objetivo de la revisión, garantizando que la evidencia sintetizada refleje específicamente las estrategias desplegadas en entornos de mensajería cifrada y su relación con la prevención de conflictos sociales.

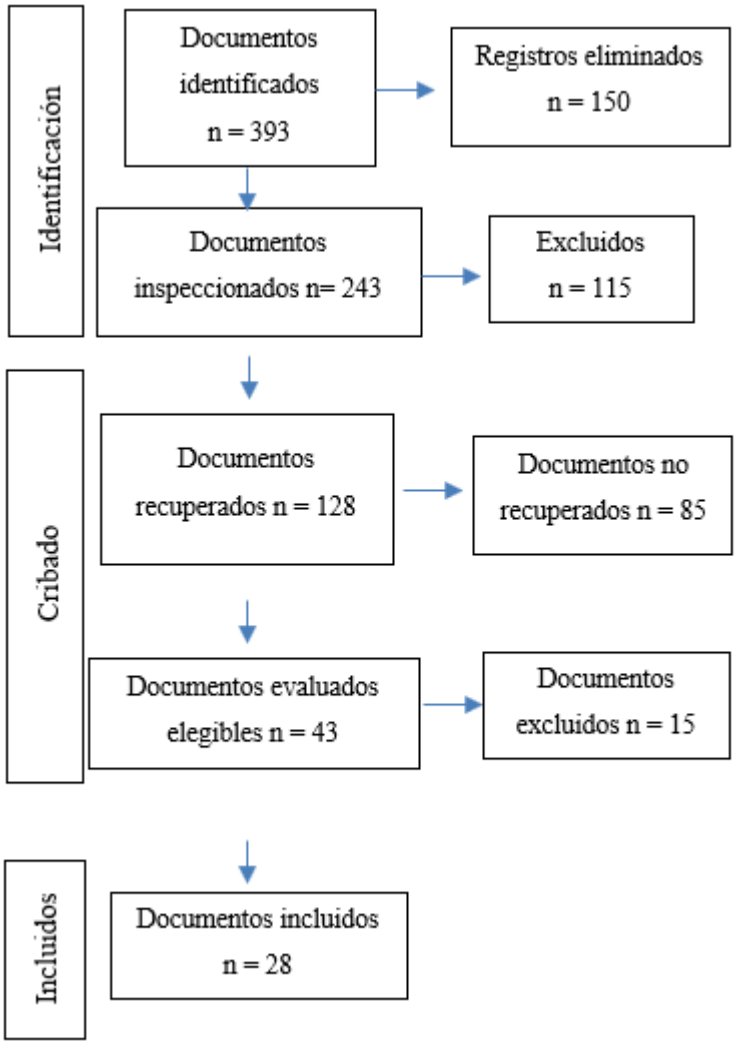

Figura 1. Identificación de estudios que utilizan el método prismático

ISSN: 2665-0398

Resultados

Autor	Plataforma	Estrategia identificada	Evidencia
Frischlich et al., (2024)	WhatsApp	Bots de verificación.	Entrevistas a 18 usuarios.
Lee & Fussell (2025)	LINE	Chatbot de fact-checking en grupos privados.	Entrevistas a 27 usuarios.
Reis et al., (2020)	WhatsApp	Fact-checking en el dispositivo mediante huellas de contenido.	Análisis de imágenes en grupos públicos de WhatsApp en Brasil e India.
Shahid et al., (2025)	WhatsApp	Moderación por administradores de grupos.	Entrevistas a 32 administradores y revisión de 30 grupos.
Watkin & Conway (2022)	Facebook, Twitter, YouTube, TikTok, Discord y Telegram	Herramientas de seguridad, control de usuarios, educación y construcción de confianza.	Análisis de 436 blogs oficiales de plataformas.

Tabla 2. Estrategias técnicas de plataformas

Los resultados evidencian que las estrategias técnicas orientadas a enfrentar la desinformación en entornos cifrados se concentran, principalmente, en herramientas de verificación automatizada, chatbots, advertencias asociadas a contenidos previamente contrastados, sistemas de huellas digitales y modalidades de moderación descentralizada. Estas medidas procuran intervenir sobre la circulación de información falsa sin vulnerar necesariamente el cifrado de extremo a extremo, lo que revela una tensión medular entre la contención de la desinformación y la preservación de la privacidad comunicacional.

Desde esta perspectiva, los estudios sobre bots verificadores, chatbots y mecanismos de rastreo técnico permiten advertir que la tecnología puede contribuir a la identificación temprana de contenidos engañosos; no obstante, su eficacia queda supeditada a factores sociotécnicos como la confianza del usuario, la inteligibilidad de la herramienta, la facilidad de uso y la oportunidad con que la corrección ingresa al mismo circuito donde se propaga el contenido falso.

Asimismo, los hallazgos muestran que la intervención técnica no recae exclusivamente en la arquitectura de la plataforma, sino también en actores internos de los grupos, especialmente en los administradores, quienes asumen funciones de moderación, filtrado y contención de interacciones potencialmente problemáticas. Ello permite sostener que, en los servicios de mensajería privada, la efectividad de las estrategias técnicas no depende únicamente del diseño algorítmico, sino de una articulación más compleja entre infraestructura tecnológica, gobernanza comunitaria y corresponsabilidad de los usuarios. Sin embargo, la evidencia disponible todavía resulta insuficiente para establecer con precisión su impacto directo en

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

la prevención de conflictos sociales, debido a que varios estudios se concentran en la aceptación, el diseño o el funcionamiento operativo de las herramientas, pero no siempre examinan sus efectos sostenidos sobre la reducción del escalamiento conflictivo**.**

Autor	País/región	Política o regulación	Evidencia
de Keulenaar & Alves dos Santos Jr. (2026)	Brasil	Gobernanza de moderación y discurso político.	Análisis digital de varias plataformas, incluido Telegram.
Dudar & Molotkina (2025)	Ucrania	Comunicación estatal y centros contra desinformación.	Análisis institucional y de casos 2022–2025.
Kabha et al., (2019)	UAE, Reino Unido e India	Regulación legal contra fake news en WhatsApp.	Revisión cualitativa comparada.
Ketners (2026)	Estados bálticos / Unión Europea	DSA, DMA y supervisión digital.	Análisis regulatorio-documental.
Kurnia et al., (2024)	Indonesia / West Java	Contranarrativas gubernamentales contra hoaxes.	Encuesta a 5,000 usuarios de Instagram y WhatsApp.
Medeiros & Singh (2020)	India / WhatsApp	Responsabilidad de intermediarios, diseño de plataforma y alfabetización mediática.	Estudio de caso sobre rumores y linchamientos.
Meyer & Vetulani-Cęgiel (2025)	Unión Europea	Transparencia en publicidad política y Código de Buenas Prácticas sobre Desinformación.	Análisis comparado de políticas de la UE y plataformas.
Oleksiyuk (2025)	Ucrania	Acceso a información oficial en guerra.	Revisión legislativa, análisis comparado y casos.
Santin et al., (2024)	Brasil	Proyecto de Ley de Fake News y responsabilidad de big techs.	Análisis jurídico-normativo.
Vijaykumar et al., (2021)	Reino Unido y Brasil / WhatsApp	Corrección institucional mediante fuente oficial.	Experimento aleatorizado con 1,454 usuarios de WhatsApp.

Tabla 3. Marcos regulatorios y políticas gubernamentales

Los resultados evidencian que las respuestas gubernamentales frente a la desinformación en plataformas digitales se han orientado, principalmente, hacia la regulación legal, la responsabilidad de los intermediarios, la transparencia, la comunicación oficial, la construcción de contranarrativas y el acceso a información pública verificable. Los estudios revisados muestran que la actuación estatal no se limita a sancionar o restringir la circulación de contenidos falsos, sino que también busca robustecer mecanismos institucionales de respuesta ante rumores, propaganda y operaciones de manipulación informativa. En contextos como India, Brasil, Ucrania y la Unión Europea, la regulación aparece imbricada con problemas de seguridad pública, violencia colectiva, propaganda política, crisis sanitaria y deterioro de la confianza institucional.

En conjunto, estos hallazgos permiten advertir que la principal tensión regulatoria se sitúa entre la protección de la seguridad pública y el resguardo de

ISSN: 2665-0398

la privacidad comunicacional, especialmente cuando la desinformación se propaga en aplicaciones cifradas. Las políticas más consistentes no son aquellas que promueven una vigilancia indiscriminada, sino las que articulan regulación proporcional, transparencia procedimental, información oficial oportuna y cooperación con plataformas tecnológicas o actores verificadores. No obstante, la evidencia también sugiere que las respuestas gubernamentales pueden resultar insuficientes cuando carecen de credibilidad pública o son percibidas como dispositivos de censura. Por ello, su efectividad no depende únicamente de la solidez normativa, sino también del grado de confianza ciudadana en las instituciones encargadas de implementarlas.

Autor	Contexto	Estrategia identificada	Evidencia
da Costa Figueiredo et al., (2023)	Brasil / WhatsApp	Alfabetización mediática digital para adultos mayores.	Estudio cuantitativo con 347 participantes.
Tirado García & Alonso-Muñoz (2024)	España / WhatsApp	Difusión de verificaciones por agencias de fact-checking.	Análisis de 258 mensajes de Newtral, Maldita y EFE Verifica.
Garimella (2022)	India / WhatsApp	Fact-checking comunitario.	Análisis mixto de grupos públicos y privados.
Martín-Neira et al., (2023)	Iberoamérica / salud y ciencia	Fact-checking periodístico.	Análisis de 240 publicaciones.
Meyrer & Kersch (2022)	Brasil / WhatsApp	Alfabetización mediática crítica en estudiantes.	Estudio de caso en clases de inglés.
Navumau et al., (2025)	Ucrania / Telegram	Coordinación ciudadana y voluntaria en guerra.	20 entrevistas y observación de canales.
Pereira Tavares et al., (2024)	Brasil / WhatsApp	Fact-checking sanitario mediante Agência Lupa.	Raspado de datos y análisis quali-cuantitativo.
Sádaba et al., (2023)	España / WhatsApp	Curso de alfabetización mediática para mayores de 50 años.	Experimento con grupo control y experimental.
Shahi & Hale (2025)	India / WhatsApp	Tiplines de fact-checking electoral.	Análisis de 580 reclamos enviados por 451 usuarios.
Talabi et al. ,(2022)	Nigeria / WhatsApp	Consejería digital contra fake news sobre vacunas.	Dos experimentos con intervención posterior.
Thu et al., (2026)	Myanmar / crisis múltiple	Verificación comunitaria e información sanitaria híbrida.	Mapeo participativo con 24 actores.
Trollip et al., (2024)	Sudáfrica / WhatsApp y Facebook	Campaña multilingüe y multimodal contra rumores.	Caso práctico en encuesta nacional de VIH.
Wang (2022)	Taiwán / Line y Facebook	Fact-checking en plataformas públicas y privadas.	Experimento con 601 participantes y encuesta nacional con 1,060 personas.

Tabla 4. Iniciativas de la sociedad civil

Los resultados muestran que la sociedad civil desempeña un papel estratégico en la mitigación de la desinformación mediante iniciativas de fact-checking, alfabetización mediática, líneas de

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

verificación, campañas comunitarias, consejería digital y redes híbridas de comunicación. A diferencia de las respuestas técnicas o estatales, estas intervenciones se distinguen por su proximidad con los usuarios, su plasticidad cultural y su capacidad para actuar dentro de comunidades donde la información falsa circula con mayor densidad relacional. Los estudios incluidos evidencian que el fact-checking alcanza mayor eficacia cuando utiliza los mismos canales por los que se propaga la desinformación, como WhatsApp, Line o Telegram, y cuando se apoya en fuentes socialmente reconocidas o dotadas de legitimidad comunitaria.

Asimismo, las intervenciones educativas y comunitarias demuestran que la prevención de la desinformación no se agota en el desmentido de contenidos, sino que exige fortalecer capacidades críticas, cultivar confianza interpersonal y adecuar los mensajes a públicos específicos, como adultos mayores, estudiantes, comunidades sanitarias o poblaciones expuestas a situaciones de crisis.

En escenarios de guerra, emergencia sanitaria o inestabilidad múltiple, las iniciativas impulsadas por la sociedad civil articulan canales digitales, medios tradicionales, comunicación presencial y verificación comunitaria, con el propósito de reducir la incertidumbre y contener la propagación de rumores. No obstante, su principal limitación radica en la dificultad para alcanzar una escala proporcional a la velocidad con que la desinformación circula en redes cerradas. Por ello, su mayor contribución se sitúa en la prevención localizada, la reconstrucción de confianza y el fortalecimiento de la resiliencia comunitaria.

Discusión de resultados

Los resultados de esta revisión sistemática evidencian que las estrategias implementadas para contrarrestar la desinformación en entornos cifrados durante el período 2015-2026 se organizan en tres grandes ámbitos de intervención: las respuestas técnicas de las plataformas, las políticas gubernamentales y regulatorias, y las iniciativas promovidas por la sociedad civil. En conjunto, los hallazgos permiten sostener que la mitigación de la desinformación en plataformas de mensajería privada no puede depender de una única estrategia, dado que estos entornos articulan privacidad comunicacional, alta confianza interpersonal, baja visibilidad pública y dificultades estructurales para la moderación centralizada.

Esta lectura converge con lo señalado por Chagas & Da-Costa (2023), quienes advierten que la opacidad de WhatsApp y otros servicios de mensajería condiciona tanto la investigación académica como la capacidad de intervención pública. Asimismo, los resultados coinciden con Gursky et al., (2022), quienes sostienen que las aplicaciones de chat operan como espacios intermedios entre la comunicación privada y la

ISSN: 2665-0398

circulación pública, permitiendo que la desinformación se planifique, se disemine y escale hacia otros ecosistemas digitales.

Respecto a las estrategias técnicas de las plataformas, los resultados muestran que las respuestas más recurrentes se relacionan con bots de verificación, chatbots de fact-checking, sistemas de advertencia basados en contenidos previamente verificados, moderación por administradores de grupos y herramientas de seguridad o control de usuarios. Esta evidencia converge parcialmente con los hallazgos de Frischlich et al., (2024), quienes identifican que los bots de verificación pueden ser percibidos como recursos útiles para contrarrestar información falsa en WhatsApp, aunque su aceptación depende de la confianza del usuario y de la forma en que la herramienta se integra en la interacción cotidiana.

De modo similar, Lee & Fussell (2025) muestran que los chatbots de verificación en grupos privados pueden apoyar la corrección de la desinformación, pero su eficacia no se explica solo por la disponibilidad técnica, sino también por la percepción de legitimidad, utilidad y baja intrusividad dentro de la conversación privada.

Estos hallazgos dialogan con Reis et al., (2020), quienes plantean que WhatsApp podría beneficiarse del uso de historias previamente verificadas para advertir a los usuarios sobre contenidos falsos sin vulnerar el cifrado de extremo a extremo. La coincidencia entre dichos aportes y los resultados de la presente revisión radica en que las soluciones técnicas más viables son aquellas que intervienen sobre metadatos, huellas de contenido, bases de datos de verificaciones o mecanismos voluntarios de consulta, antes que sobre la inspección directa de mensajes privados.

Sin embargo, la revisión también evidencia una limitación relevante: muchas estrategias técnicas se evalúan en términos de aceptación, diseño o factibilidad, pero no necesariamente en función de su impacto real sobre la reducción de rumores, discursos polarizantes o procesos de escalamiento conflictivo. Esta brecha coincide con la advertencia de Melo et al., (2024), quienes muestran que incluso medidas implementadas por WhatsApp, como las etiquetas de reenvío, pueden ser eludidas o aplicarse de manera incompleta, lo que restringe su capacidad para contener cadenas de desinformación.

La moderación descentralizada por administradores de grupos constituye otro hallazgo relevante. Shahid et al., (2025) muestran que la regulación de contenidos problemáticos en grupos de WhatsApp no responde a un modelo homogéneo, sino a prácticas situadas, dependientes de normas comunitarias, vínculos de confianza y criterios informales de administración. Este resultado amplía la comprensión tradicional de la moderación, pues indica que, en entornos cifrados, la gobernanza de la

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

desinformación no se ubica exclusivamente en la arquitectura de la plataforma, sino también en actores internos de los grupos.

Esta evidencia converge con Pasquetto et al., (2022), quienes sostienen que los lazos fuertes y las relaciones de pertenencia pueden favorecer la redistribución de desmentidos en WhatsApp. No obstante, también se advierte una tensión significativa: los mismos vínculos de confianza que facilitan la corrección pueden incrementar la credibilidad de mensajes falsos cuando estos provienen de familiares, amistades o comunidades ideológicamente afines.

En cuanto a los marcos regulatorios y las políticas gubernamentales, los resultados evidencian que las estrategias estatales se han orientado hacia la responsabilidad de intermediarios, la regulación de noticias falsas, la transparencia en publicidad política, la comunicación oficial, las contranarrativas y el acceso a información pública confiable.

Este resultado coincide con Medeiros & Singh (2020), quienes sostienen que el abordaje de la desinformación en WhatsApp requiere una combinación entre responsabilidad de intermediarios, modificaciones en el diseño de las plataformas y alfabetización mediática. Asimismo, Kabha et al., (2019) muestran que países como Emiratos Árabes Unidos, Reino Unido e India han ensayado respuestas legales frente a la desinformación en WhatsApp, aunque con diferencias sustantivas en el equilibrio entre control estatal, libertad de expresión y privacidad.

La revisión permite advertir que la principal tensión de las políticas gubernamentales se sitúa entre la seguridad pública y la privacidad de las comunicaciones. Este hallazgo converge con Santin et al., (2024), quienes analizan el debate brasileño sobre regulación de noticias falsas y responsabilidad de las grandes plataformas, mostrando que la intervención estatal puede ser necesaria ante daños sociales concretos, pero también riesgosa si deriva en censura, vigilancia o control excesivo del debate público.

En la misma línea, Meyer & Vetulani-Cęgiel (2025) evidencian que las iniciativas europeas centradas en transparencia política y publicidad digital intentan desplazar la regulación desde la simple remoción de contenidos hacia la identificación de actores, financiamiento y mecanismos de circulación. Este énfasis resulta particularmente relevante para los entornos cifrados, donde la eliminación directa de mensajes es limitada y donde la transparencia sobre redes de influencia puede resultar más viable que la inspección masiva de comunicaciones.

Los estudios desarrollados en contextos de guerra o crisis institucional refuerzan la importancia de la comunicación oficial como estrategia de prevención. Dudar & Molotkina (2025) y Oleksiyuk

ISSN: 2665-0398

(2025) muestran que, en el caso ucraniano, la disponibilidad de información pública, la comunicación de crisis y el acceso a fuentes oficiales pueden fortalecer la resiliencia social frente a rumores, propaganda y manipulación informativa.

Estos hallazgos convergen con Kurnia et al., (2024), quienes evidencian que las contranarrativas gubernamentales pueden influir en las respuestas emocionales de los usuarios ante contenidos falsos, especialmente cuando las fuentes son percibidas como creíbles. También se relacionan con Vijaykumar et al., (2021), quienes muestran que la información correctiva atribuida a la Organización Mundial de la Salud incrementa la credibilidad percibida y la intención de compartir información correcta entre usuarios de WhatsApp en Reino Unido y Brasil. Sin embargo, la efectividad de estas estrategias depende de la legitimidad institucional, pues en contextos de baja confianza pública la comunicación oficial puede ser recibida con escepticismo o interpretada como propaganda.

Respecto a las iniciativas de la sociedad civil, los resultados muestran que las intervenciones más frecuentes se concentran en fact-checking, alfabetización mediática, líneas de verificación, campañas comunitarias, consejería digital y redes híbridas de comunicación. Estos hallazgos convergen con Tirado García & Alonso-Muñoz (2024), quienes evidencian que las agencias de verificación españolas utilizan canales de WhatsApp para distribuir contenidos contrastados, adaptando el fact-checking al ecosistema donde circulan los rumores.

De modo similar, Shahi & Hale (2025) muestran que las líneas de verificación de WhatsApp pueden servir como mecanismos de recepción, clasificación y respuesta ante afirmaciones dudosas en contextos electorales multilingües. La coincidencia central radica en que la sociedad civil resulta particularmente efectiva cuando interviene en los mismos canales, lenguajes y formatos donde se propaga la desinformación.

La alfabetización mediática aparece como una estrategia con evidencia significativa de efectividad, especialmente en públicos vulnerables o expuestos a información de baja calidad. Sádaba et al., (2023) demuestran que una intervención formativa puede mejorar la capacidad de los adultos mayores para detectar desinformación política, mientras que da Costa Figueiredo et al., (2023) encuentran resultados positivos en programas de alfabetización mediática digital dirigidos a personas mayores en Brasil.

Meyrer & Kersch (2022), por su parte, muestran que los estudiantes de secundaria pueden desarrollar habilidades críticas para verificar información sobre la COVID-19 cuando la intervención educativa se diseña de manera contextualizada. Estos resultados coinciden con

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

Vartiainen et al., (2023), quienes advierten que niñas, niños y adolescentes reconocen la existencia de noticias falsas, pero suelen emplear estrategias superficiales de evaluación, por lo que requieren enfoques pedagógicos más profundos sobre evidencia, manipulación, perfilamiento algorítmico y circulación digital.

El fact-checking comunitario y la corrección social constituyen otra línea relevante. Garimella (2022) muestra que los usuarios pueden participar en procesos de corrección dentro de grupos de WhatsApp, aunque la efectividad depende de quién corrige, a quién se corrige y bajo qué condiciones relacionales ocurre la intervención. Este resultado converge con Badrinathan & Chauchard (2023), quienes sostienen que las correcciones sociales pueden ser relevantes en India, pero no siempre logran reducir la creencia en información falsa cuando entran en conflicto con identidades políticas o vínculos comunitarios.

En esa misma dirección, Wang (2022) evidencia que la efectividad del fact-checking varía entre plataformas públicas y privadas, pues en espacios privados los usuarios pueden seleccionar verificaciones congruentes con sus preferencias previas. Esta divergencia muestra que el fact-checking no debe entenderse como una intervención universalmente efectiva, sino como una estrategia condicionada por la arquitectura de la plataforma, la afinidad ideológica, la confianza interpersonal y el contexto sociopolítico.

Las experiencias de la sociedad civil en contextos de crisis muestran que la mitigación de la desinformación no se reduce a corregir afirmaciones falsas, sino que implica sostener circuitos de confianza, coordinación y resiliencia comunitaria. Navumau et al., (2025) evidencian que Telegram puede funcionar como infraestructura de coordinación en contextos bélicos, al permitir la organización de ayuda, la verificación de solicitudes y el sostenimiento de redes voluntarias.

Thu et al., (2026) muestran que, en una situación de crisis múltiple en Myanmar, la difusión de información sanitaria confiable requirió una combinación de Telegram, Viber, VPN, radio, comunicación presencial y verificación comunitaria. Trollip et al., (2024), en Sudáfrica, muestran que una campaña multilingüe y multimodal permitió enfrentar rumores difundidos por WhatsApp y Facebook, así como recuperar la confianza en una encuesta nacional sobre VIH. Estos hallazgos permiten afirmar que la prevención de conflictos sociales no depende únicamente de desmentidos, sino de la capacidad de construir confianza situada, coordinar actores locales y ofrecer información útil en escenarios de incertidumbre.

En relación con el objetivo general de esta revisión, los resultados permiten sostener que las

ISSN: 2665-0398

estrategias analizadas contribuyen de manera indirecta y diferencial a la prevención de conflictos sociales. Las estrategias técnicas pueden reducir la exposición o circulación de contenidos falsos, pero su impacto es limitado si no se acompañan de confianza y adopción por parte de los usuarios.

Las políticas gubernamentales pueden generar marcos de responsabilidad, transparencia y comunicación oficial, pero su efectividad depende de su proporcionalidad y legitimidad democrática. Las iniciativas de la sociedad civil muestran mayor proximidad con las comunidades y mayor capacidad de adaptación cultural, aunque enfrentan problemas de escala, sostenibilidad y alcance frente a la velocidad de la desinformación en redes cerradas. En consecuencia, la prevención del escalamiento conflictivo exige modelos de intervención híbridos, colaborativos y sensibles al contexto.

Una primera limitación metodológica de esta revisión se relaciona con la selección de la base de datos. Al haberse empleado Scopus como fuente principal, el corpus garantiza un estándar de calidad académica; sin embargo, puede excluir literatura relevante indexada en Web of Science, SciELO, Redalyc, Dialnet, Latindex u otras bases regionales, especialmente si se considera que la desinformación en mensajería privada tiene fuerte presencia en América Latina, Asia y África. Esta limitación podría afectar la representatividad geográfica de los hallazgos y reducir la visibilidad de investigaciones publicadas en revistas locales o en idiomas distintos del inglés.

Una segunda limitación se vincula con la heterogeneidad metodológica de los estudios incluidos. El corpus integra experimentos, entrevistas, estudios de caso, análisis documental, análisis normativos, mapeos participativos, encuestas y estudios cualitativos. Esta diversidad enriquece la comprensión del fenómeno, pero dificulta la comparación directa de los niveles de efectividad.

Mientras algunos trabajos miden cambios en percepción, intención de compartir o capacidad de identificación de contenidos falsos, otros se concentran en diseño, aceptación, regulación o descripción de prácticas. Por ello, los resultados deben interpretarse como una síntesis analítica de evidencias heterogéneas, no como una medición uniforme del impacto causal de cada estrategia sobre la prevención de conflictos sociales.

Una tercera limitación está asociada con la propia naturaleza de los entornos cifrados. WhatsApp, Signal, Telegram y otras aplicaciones presentan restricciones de acceso a datos, opacidad en la circulación de mensajes y dificultades éticas para observar interacciones privadas. Esta condición limita la capacidad de los estudios para reconstruir cadenas completas de difusión, medir efectos longitudinales o determinar con precisión si

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

una intervención redujo efectivamente la escalada de conflictos.

Como advierten Chagas & Da-Costa (2023) y Martínez et al., (2022), la mensajería privada opera como una "caja negra" metodológica, por lo que muchas investigaciones dependen de grupos públicos, autoinformes, simulaciones, experimentos controlados o evidencias parciales.

Una cuarta limitación se relaciona con la noción de prevención de conflictos sociales. Aunque el objetivo del estudio se orienta a evaluar la efectividad de las estrategias frente al escalamiento conflictivo, gran parte de la literatura disponible no mide directamente variables de conflicto, violencia colectiva o polarización sostenida. En varios casos, la efectividad se infiere a partir de indicadores indirectos, como reducción de creencias falsas, mejora en alfabetización mediática, incremento de credibilidad de información correcta, mayor intención de compartir verificaciones o fortalecimiento de redes comunitarias. Esta situación exige interpretar con cautela el vínculo entre mitigación de la desinformación y prevención de conflictos, pues no toda reducción de contenidos falsos se traduce necesariamente en disminución de conflictividad social.

A partir de estas limitaciones, futuras investigaciones deberían ampliar las bases de datos consultadas e incorporar literatura indexada en repositorios regionales, especialmente en América Latina, África y Asia, donde WhatsApp, Telegram y Line cumplen funciones políticas, comunitarias y sanitarias relevantes. También sería pertinente desarrollar estudios comparativos entre países con distintos niveles de confianza institucional, regulación digital y penetración de mensajería privada, a fin de explicar por qué ciertas estrategias funcionan en algunos contextos y fracasan en otros.

Asimismo, se requieren investigaciones longitudinales que permitan evaluar el impacto sostenido de bots, chatbots, líneas de verificación, campañas de alfabetización mediática, contranarrativas y políticas regulatorias sobre la circulación de desinformación y la evolución de conflictos sociales. Estos estudios deberían incorporar indicadores más robustos de efectividad, tales como cambios en patrones de reenvío, reducción de rumores verificables, disminución de discursos hostiles, fortalecimiento de confianza comunitaria o contención de episodios de movilización conflictiva asociados a información falsa.

Otra línea futura consiste en profundizar el análisis ético y técnico de intervenciones compatibles con el cifrado de extremo a extremo. Resulta necesario explorar modelos de verificación en el dispositivo, sistemas voluntarios de consulta, advertencias no intrusivas, herramientas de alfabetización integradas en la plataforma y

ISSN: 2665-0398

mecanismos de trazabilidad respetuosos de la privacidad. Esta agenda permitiría avanzar hacia soluciones que no reproduzcan la falsa dicotomía entre seguridad pública y privacidad, sino que articulen mitigación de daños, derechos digitales y gobernanza democrática.

Finalmente, futuras investigaciones deberían prestar mayor atención a las iniciativas comunitarias y de la sociedad civil, particularmente en contextos de crisis, conflicto armado, desastres, elecciones polarizadas o emergencias sanitarias. Los hallazgos de esta revisión sugieren que la confianza interpersonal, la adaptación lingüística, la legitimidad local y la comunicación híbrida son factores decisivos para contener rumores y reducir incertidumbre. Por ello, el estudio de la desinformación en entornos cifrados debe avanzar desde enfoques centrados únicamente en plataformas o regulaciones hacia modelos ecosistémicos que integren tecnología, instituciones, comunidades y prácticas sociales de verificación.

Conclusiones

La presente revisión sistemática permitió identificar que las respuestas orientadas a contrarrestar la desinformación en entornos cifrados se agrupan en tres dimensiones principales: estrategias técnicas de las plataformas, políticas gubernamentales y regulatorias, e iniciativas promovidas por la sociedad civil. Los hallazgos evidencian que las plataformas de mensajería privada han desarrollado o evaluado mecanismos como bots de verificación, chatbots de fact-checking, sistemas de advertencia basados en contenidos previamente contrastados, huellas digitales de contenido y modalidades de moderación descentralizada. No obstante, su efectividad se encuentra condicionada por la confianza de los usuarios, la compatibilidad con la privacidad comunicacional y la capacidad de intervenir oportunamente en los circuitos donde circula la información falsa. En paralelo, las respuestas gubernamentales se orientan hacia la regulación legal, la responsabilidad de intermediarios, la transparencia, la comunicación oficial y las contranarrativas institucionales; mientras que las iniciativas de la sociedad civil destacan por su proximidad comunitaria, alfabetización mediática, líneas de verificación, campañas multilingües y estrategias híbridas de comunicación. En conjunto, estos resultados contribuyen al campo de estudio al mostrar que la mitigación de la desinformación en servicios de mensajería privada exige respuestas integradas, contextuales y multisectoriales, debido a que ninguna intervención aislada resulta suficiente para enfrentar un fenómeno atravesado por la privacidad, la confianza interpersonal, la polarización y la circulación vertiginosa de contenidos falsos.

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

En relación con el objetivo de evaluar las estrategias implementadas por plataformas de mensajería privada, gobiernos y sociedad civil para contrarrestar la desinformación en entornos cifrados, así como su efectividad en la prevención de conflictos sociales entre 2015 y 2026, los resultados permiten concluir que dicha efectividad es diferenciada, contingente e indirecta. Las respuestas técnicas contribuyen a reducir la exposición o redistribución de contenidos falsos; sin embargo, presentan limitaciones cuando dependen exclusivamente de la iniciativa del usuario o cuando no logran insertarse en los flujos comunicativos privados. Las políticas gubernamentales pueden fortalecer la rendición de cuentas, la transparencia y la circulación de información oficial, aunque su impacto depende de la legitimidad democrática, la proporcionalidad normativa y la capacidad de resguardar la privacidad de las comunicaciones. Por su parte, las iniciativas de la sociedad civil muestran mayor potencial preventivo en el plano comunitario, especialmente cuando combinan verificación de hechos, educación crítica, confianza interpersonal y adaptación lingüística o cultural. No obstante, la prevención de conflictos sociales no aparece como un efecto directo ni plenamente medido en todos los estudios revisados, sino como una consecuencia probable derivada de la reducción de rumores, el fortalecimiento de la alfabetización mediática, la confianza pública y la contención de narrativas polarizantes.

El estudio corresponde a una revisión sistemática de la literatura científica, desarrollada mediante una estrategia de búsqueda estructurada y orientada por preguntas de investigación vinculadas con las acciones de plataformas, gobiernos y sociedad civil. Este enfoque metodológico permitió organizar la evidencia disponible, comparar modalidades de intervención y reconocer patrones comunes en distintos contextos geográficos, políticos y comunicacionales. Al tratarse de una revisión sistemática, las conclusiones no se sustentan en la producción de datos primarios, sino en la síntesis crítica de investigaciones previamente publicadas en revistas científicas. Por ello, su principal aporte radica en ofrecer una lectura integrada del estado de la investigación sobre desinformación en mensajería privada y entornos cifrados, destacando tanto las convergencias analíticas como las brechas aún persistentes en la literatura especializada.

Como reflexión final, los hallazgos sugieren que la lucha contra la desinformación en entornos cifrados debe avanzar hacia modelos de gobernanza colaborativa que articulen innovación tecnológica, garantías de privacidad, regulación proporcional, alfabetización mediática y participación comunitaria. Futuras investigaciones deberían profundizar en estudios longitudinales que midan el impacto sostenido de las intervenciones sobre indicadores concretos de conflictividad social, tales como polarización, discursos hostiles, violencia

ISSN: 2665-0398

colectiva, desconfianza institucional o movilización basada en rumores. Asimismo, resulta necesario ampliar la evidencia empírica en regiones del Sur Global, donde WhatsApp, Telegram, Line y otras aplicaciones de mensajería cumplen funciones políticas, sanitarias y comunitarias especialmente relevantes. Finalmente, se recomienda explorar diseños de intervención compatibles con el cifrado de extremo a extremo, capaces de preservar los derechos digitales sin renunciar a la prevención de daños sociales derivados de la desinformación.

Referencias

Badrinathan, S., & Chauchard, S. (2023). "I don't think that's true, bro!" Social corrections of misinformation in India. *The International Journal of Press/Politics*. Documento en línea. Disponible https://doi.org/10.1177/19401612231158770

Chagas, V., & Da-Costa, G. (2023). WhatsApp and transparency: An analysis on the effects of digital platforms' opacity in political communication research agendas in Brazil. *Profesional de la Información, 32*(2). Documento en línea. Disponible https://doi.org/10.3145/epi.2023.mar.23

Cocha, D. P. V., Chicaiza, F. P. V., Guevara, A. M. G., Palacios, I. A. M., & Arce, M. (2024). Desinformación en la era digital: El papel de las redes sociales en la propagación de noticias falsas durante conflictos globales. *Latam Revista Latinoamericana de Ciencias Sociales y Humanidades, 5*(2). Documento en línea. Disponible https://doi.org/10.56712/latam.v5i2.1865

da Costa Figueiredo, C., Antonioli, M. E., & Guimarães Gil, P. (2023). A efetividade de um programa de alfabetização em mídia digital para idosos brasileiros. *Comunicação, Mídia e Consumo, 20*(58), 219–241. Documento en línea. Disponible https://doi.org/10.18568/CMC.V20I58.2792

de Keulenaar, E., & Alves dos Santos Jr., M. (2026). Normative dislocation: When platforms moderate without memory. *New Media & Society, 28*(4), 1437–1463. Documento en línea. Disponible https://doi.org/10.1177/14614448251364814

Díez-Garrido, M., Farpón, C. R., & Cano-Orón, L. (2021). Desinformación en las redes de mensajería instantánea: Estudio de las fake news en los canales relacionados con la ultraderecha española en Telegram. *Miguel Hernández Communication Journal, 12*, 467–489. Documento en línea. Disponible https://doi.org/10.21134/mhjournal.v12i.1292

Duarte, J. M. S., & Rosa, R. M. (2023). Desinformación. *Eunomía. Revista en Cultura de la Legalidad, 24*, 236–249. Documento en línea. Disponible https://doi.org/10.20318/eunomia.2023.7663

Dudar, V., & Molotkina, Y. (2025). Social communication in crisis situations: The case of Ukrainian media, 2022–2025. *Society. Document. Communication, 10*(3), 66–81. Documento en línea. Disponible https://doi.org/10.69587/sdc/3.2025.66

Frischlich, L., Klapproth, J., Frank, S., Heckmann, M., Kunze, S. E., & Murgas, T. (2024). Fighting fakes on WhatsApp: Audience perspectives on fact bots as countermeasures. *Digital Journalism, 12*(5), 700–720. Documento en línea. Disponible https://doi.org/10.1080/21670811.2024.2341299

Garimella, K. (2022). Community-driven fact-checking on WhatsApp: Who fact-checks whom, why, and with what effect? *Association for the Advancement of Artificial Intelligence*.

Gursky, J., Riedl, M., Joseff, K., & Woolley, S. (2022). Chat apps and cascade logic: A multi-platform perspective on India, Mexico, and the United States. *Social Media + Society, 8*(2), 1–

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

15. Documento en línea. Disponible https://doi.org/10.1177/20563051221094773

Kabha, R., Kamel, A., Elbahi, M., & Narula, S. (2019). Comparison study between the UAE, the UK, and India in dealing with WhatsApp fake news. *Journal of Content, Community & Communication, 10*(5), 176–185. Documento en línea. Disponible https://doi.org/10.31620/JCCC.12.19/18

Ketners, K. (2026). Social platforms and the transformation of media landscapes in the Baltic states. *Society. Document. Communication, 11*(1), 38–49. Documento en línea. Disponible https://doi.org/10.69587/sdc/1.2026.38

Krishnan, N., Gu, J., Tromble, R., & Abroms, L. C. (2021). Examining how various social media platforms have responded to COVID-19 misinformation. *HKS Misinformation Review*. Documento en línea. Disponible https://doi.org/10.37016/mr-2020-85

Kurnia, S. S., Rahman, Z., Cakranegar, D. I., Abdulla, S. I., Setiawan, D. A., Agustini, P. M., & Yenrizal. (2024). Effect of counter-narratives and credibility of sources on emotional response: A study of Instagram and WhatsApp followers. *Journal of Intercultural Communication, 24*(1), 161–173. Documento en línea. Disponible https://doi.org/10.36923/jicc.v24i1.170

Lee, T.-H., & Fussell, S. R. (2025). Countering misinformation in private messaging groups: Insights from a fact-checking chatbot. *Proceedings of the ACM on Human-Computer Interaction, 9*(GROUP), Article GROUP10. Documento en línea. Disponible https://doi.org/10.1145/3701189

Martínez, I. C., Martínez, J. M. R., Zapata, C. M., & Moreno, S. M. (2022). Conversación y difusión de información política en WhatsApp: Un análisis de la "caja negra" desde las teorías de la interacción en redes sociales. *Revista de Comunicación, 21*(1), 117–136. Documento en línea. Disponible https://doi.org/10.26441/rc21.1-2022-a6

Martín-Neira, J.-I., Trillo-Domínguez, M., & Olvera-Lobo, M.-D. (2023). Ibero-American journalism in the face of scientific disinformation: Fact-checkers' initiatives on the social network Instagram. *Profesional de la Información, 32*(5), Article e320503. Documento en línea. Disponible https://doi.org/10.3145/epi.2023.sep.03

Medeiros, B., & Singh, P. (2020). Addressing misinformation on WhatsApp in India through intermediary liability policy, platform design modification, and media literacy. *Journal of Information Policy, 10*, 276–298. Documento en línea. Disponible https://doi.org/10.5325/jinfopoli.10.2020.0276

Melo, P., Hoseini, M., Zannettou, S., & Benevenuto, F. (2024). Don't break the chain: Measuring message forwarding on WhatsApp. *Proceedings of the International AAAI Conference on Web and Social Media, 18*, 1054–1067. Documento en línea. Disponible https://doi.org/10.1609/icwsm.v18i1.31372

Meyer, T., & Vetulani-Cęgiel, A. (2025). Transparency as an empty signifier? Assessing transparency in EU and platform initiatives on online political advertising and actors. *Policy & Internet, 17*, Article e417. Documento en línea. Disponible https://doi.org/10.1002/poi3.417

Meyrer, K. P., & Kersch, D. F. (2022). Can high school students check the veracity of information about COVID-19? A case study on critical media literacy in Brazilian ESL classes. *Journal of Media Literacy Education, 14*(1), 14–28. Documento en línea. Disponible https://doi.org/10.23860/JMLE-2022-14-1-2

Navumau, V., Matveieva, O., Aal, K., Wulf, V., & Rohde, M. (2025). Telegram as wartime infrastructure: Alternative supply chains in Dnipropetrovsk region, winter 2025. *I-com, 24*(3), 531–543. Documento en línea. Disponible https://doi.org/10.1515/icom-2025-0044

Oleksiyuk, T. (2025). The right to access official information as a resilience-improving tool: Ukrainian lessons during wartime. *Social*

Sciences & Humanities Open, 11, Article 101549. Documento en línea. Disponible https://doi.org/10.1016/j.ssaho.2025.101549

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

Pereira Tavares, L., Sá Brasileiro, F., & Costa de Brito, H. (2024). Desinformação em saúde na pós-pandemia: Uma análise a partir da plataforma de fact-checking Agência Lupa. *Encontros Bibli, 29*, Article e98810. Documento en línea. Disponible https://doi.org/10.5007/1518-2924.2024.e98810

Reis, J. C. S., Melo, P., Garimella, K., & Benevenuto, F. (2020). Can WhatsApp benefit from debunked fact-checked stories to reduce misinformation? *Harvard Kennedy School Misinformation Review, 1*(5). Documento en línea. Disponible https://doi.org/10.37016/mr-2020-035

Sádaba, C., Salaverría, R., & Bringué, X. (2023). Overcoming the age barrier: Improving older adults' detection of political disinformation with media literacy. *Media and Communication, 11*(4), 113–123. Documento en línea. Disponible https://doi.org/10.17645/mac.v11i4.7090

Santin, J. R., Dai Pra, M., & Faccini Neto, O. (2024). Como regular as fake news no Brasil: Análise do Projeto de Lei n. 2630, que institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. *Seqüência: Estudos Jurídicos e Políticos, 45*(97), Article e98509. Documento en línea. Disponible https://doi.org/10.5007/2177-7055.2024.e98509

Shahi, G. K., & Hale, S. A. (2025). WhatsApp tiplines and multilingual claims in the 2021 Indian assembly elections. *Online Social Networks and Media, 49*, Article 100323. Documento en línea. Disponible https://doi.org/10.1016/j.osnem.2025.100323

Shahid, F., Agarwal, D., & Vashistha, A. (2025). One style does not regulate all: Moderation practices in public and private WhatsApp groups. *Proceedings of the ACM on Human-Computer Interaction, 9*(2), Article CSCW144. Documento en línea. Disponible https://doi.org/10.1145/3711042

Talabi, F. O., Ugbor, I. P., Talabi, M. J., Ugwuoke, J. C., Oloyede, D., Aiyesimoju, A. B., & Ikechukwu-Ilomuanya, A. B. (2022). Effect of a social media-based counselling intervention in countering fake news on COVID-19 vaccine in Nigeria. *Health Promotion International, 37*(2), Article daab140. Documento en línea. Disponible https://doi.org/10.1093/heapro/daab140

Thu, H., Thu, M. M., & Ali, S. H. (2026). "The first person to reach you becomes dearest to you": Mapping health information diffusion in Myanmar's polycrisis through digital strategies, networked pathways, and innovative solutions. *Social Science & Medicine, 399*, Article 119241. Documento en línea. Disponible https://doi.org/10.1016/j.socscimed.2026.119241

Tirado García, A., & Alonso-Muñoz, L. (2024). Estrategia de difusión de las agencias de verificación españolas en sus canales de WhatsApp. *Index.comunicación, 14*(2), 33–56. Documento en línea. Disponible https://doi.org/10.62008/ixc/14/02Estrat

Trollip, K., Gastrow, M., Ramlagan, S., & Shean, Y. (2024). Harnessing multimodal and multilingual science communication to combat misinformation in a diverse country setting. *Journal of Science Communication, 23*(9), Article N01. Documento en línea. Disponible https://doi.org/10.22323/2.23090801

Vijaykumar, S., Jin, Y., Rogerson, D., Lu, X., Sharma, S., Maughan, A., Fadel, B., Silva de Oliveira Costa, M., Pagliari, C., & Morris, D.

(2021). How shades of truth and age affect responses to COVID-19 (mis)information: Randomized survey experiment among WhatsApp users in UK and Brazil. *Humanities and Social Sciences Communications, 8*, Article 88. Documento en línea. Disponible https://doi.org/10.1057/s41599-021-00752-7

Wang, A. H.-E. (2022). PM me the truth? The conditional effectiveness of fact-checks across social media sites. *Social Media + Society, 8*(2), 1–15. Documento en línea. Disponible https://doi.org/10.1177/20563051221098347

Watkin, A.-L., & Conway, M. (2022). Building social capital to counter polarization and extremism? A comparative analysis of tech platforms' official blog posts. *First Monday, 27*(5). Documento en línea. Disponible https://doi.org/10.5210/fm.v27i5.12611

Vartiainen, H., Kahila, J., Tedre, M., Sointu, E., & Valtonen, T. (2023). More than fabricated news reports: Children's perspectives and experiences of fake news. *Journal of Media Literacy Education, 15*(2), 17–30. https://doi.org/10.23860/JMLE-2023-15-2-2