

Desinformación, sesgos cognitivos e inteligencia artificial generativa: el modelo ABCD (Analizar, Buscar, Contrastar y Dudar) para la verificación crítica de la información

Disinformation, cognitive biases and generative artificial intelligence: the ABCD Model (Analyse, Search, Contrast and Doubt) for critical information verification

Nuria San-Servando-Hernández

Citación recomendada:

San-Servando-Hernández, Nuria (2026). "Desinformación, sesgos cognitivos e inteligencia artificial generativa: el modelo ABCD (Analizar, Buscar, Contrastar y Dudar) para la verificación crítica de la información [Disinformation, cognitive biases and generative artificial intelligence: the ABCD Model (Analyse, Search, Contrast and Doubt) for critical information verification]". *Infonomy*, 4(5) e26033.

<https://doi.org/10.3145/infonomy.26.033>

Artículo recibido: 17-07-2026

Artículo aprobado: 16-08-2026



Nuria San-Servando-Hernández

<https://orcid.org/0009-0002-5626-4972>

<https://directorioexit.info/ficha1390>

Consultora

Santa Eulalia, 23

26580 Arnedo (La Rioja)

nsanservando@yahoo.es

Resumen

La transformación digital ha modificado la producción, circulación e interpretación de la información, situando a la desinformación entre los principales riesgos para las sociedades democráticas. Este trabajo analiza las narrativas que sustentan los procesos de desinformación en los medios digitales desde una perspectiva interdisciplinar que integra la comunicación, la psicología cognitiva, la ciencia de la información y la inteligencia artificial (IA). A



partir de la revisión de la bibliografía científica reciente y del análisis de cinco casos representativos —*Cambridge Analytica*, la "infodemia" de la COVID-19, los disturbios de Southport, el *Pizzagate* y el deepfake de Zelenski—, se examina cómo la eficacia de la manipulación informativa depende tanto de la arquitectura narrativa del mensaje, de los sesgos cognitivos del receptor y de la lógica algorítmica de las plataformas. En este escenario se propone un modelo de análisis del discurso orientado a la verificación crítica, denominado Modelo ABCD (Analizar, Buscar, Contrastar y Dudar), que integra los elementos clásicos del proceso comunicativo con la evaluación de los sesgos cognitivos y de la arquitectura algorítmica. Se concluye que la transición de la sociedad de la información hacia una sociedad de la verificación exige una alfabetización mediática crítica adaptada a la era de la IA generativa.

Palabras clave

Desinformación; Infodemia; Alfabetización mediática; Sesgos cognitivos; Pensamiento crítico; Inteligencia artificial generativa; *Deepfakes*; Análisis del discurso; Verificación de la información; *Fact-checking*; *Framing*; Ecosistema digital.

Abstract

Digital transformation has reshaped how information is produced, circulated and interpreted, placing disinformation among the main risks to democratic societies. This paper examines the narratives underlying disinformation processes in digital media from an interdisciplinary perspective that integrates communication studies, cognitive psychology, information science and artificial intelligence (AI) research. Building on a review of recent scientific literature and the analysis of five representative cases — *Cambridge Analytica*, the COVID-19 infodemic, the Southport riots, *Pizzagate* and the Zelensky deepfake—, the study examines how the effectiveness of information manipulation depends on the narrative architecture of the message, the cognitive biases of the receiver, and the algorithmic logic of digital platforms. Based on these findings, the paper proposes a discourse-analysis model for critical verification, the ABCD Model (Analyse, Búsqueda/Search, Contrast, and Doubt), which integrates the classic elements of the communicative process with the evaluation of cognitive biases and algorithmic architecture. The paper concludes that the shift from an information society towards a verification society requires a critical media literacy adapted to the era of generative AI.

Keywords

Disinformation; Infodemic; Media literacy; Cognitive biases; Critical thinking; Generative artificial intelligence; *Deepfakes*; Discourse analysis; Information Verification; *Fact-checking*; *Framing*; Digital ecosystem.

1. Introducción

La transformación digital ha modificado profundamente los procesos de producción, distribución y consumo de información, configurando un nuevo modelo comunicativo caracterizado por la hiperconectividad, la sobreabundancia de contenidos y la inmediatez en la circulación de mensajes. Internet y, especialmente, las plataformas digitales y las redes sociales, han democratizado el acceso a la información y la posibili-

dad de generar contenidos, desplazando el tradicional papel de los medios de comunicación como principales intermediarios del discurso público (**Castells**, 2009; **Van Dijck et al.**, 2018). Sin embargo, esta apertura del espacio comunicativo ha venido acompañada de importantes desafíos relacionados con la calidad de la información, la veracidad de los contenidos y la confianza en las instituciones informativas.

En este nuevo escenario, la desinformación ha dejado de constituir un fenómeno marginal para convertirse en uno de los principales riesgos para las sociedades democráticas. Organismos internacionales como el *Foro Económico Mundial (World Economic Forum, 2024)* sitúan la desinformación y la información manipulada entre las amenazas globales más relevantes a corto plazo debido a su capacidad para influir en procesos electorales, generar polarización social, deteriorar la confianza institucional y afectar a la salud pública. La pandemia provocada por la COVID-19 evidenció especialmente esta problemática, dando lugar al concepto de "infodemia", definido por la *Organización Mundial de la Salud* como la rápida propagación de un volumen excesivo de información —tanto veraz como falsa— que dificulta a la ciudadanía identificar fuentes fiables y adoptar decisiones fundamentadas en información contrastada.

La inteligencia artificial (IA), particularmente el desarrollo de modelos generativos capaces de producir textos, imágenes, vídeos y audios con un elevado grado de realismo, ha incrementado la complejidad del ecosistema informativo. Herramientas accesibles para cualquier usuario permiten actualmente crear contenidos ficticios prácticamente indistinguibles de los originales, ampliando las posibilidades de manipulación mediante *deepfakes*, imágenes hiperrealistas o campañas automatizadas de desinformación. No obstante, atribuir la problemática exclusivamente a la tecnología supondría simplificar un fenómeno mucho más complejo: la IA constituye una herramienta de enorme potencial para el desarrollo científico, sanitario, educativo o empresarial, y el verdadero desafío reside en el uso social que se hace de estas tecnologías y en la capacidad de la ciudadanía para interpretar críticamente la información que consume.

Numerosos estudios han demostrado que la difusión de información falsa no responde únicamente a factores tecnológicos, sino también a mecanismos psicológicos y sociales que condicionan la forma en que las personas procesan la información. Los sesgos cognitivos descritos por **Kahneman** (2012), como el sesgo de confirmación, la heurística de disponibilidad o el efecto halo, favorecen la aceptación y difusión de contenidos que refuerzan creencias previas o generan respuestas emocionales intensas. Del mismo modo, los algoritmos de recomendación empleados por las principales plataformas digitales priorizan contenidos capaces de maximizar la interacción del usuario, favoreciendo la formación de cámaras de eco y procesos de polarización que reducen la diversidad informativa disponible para cada usuario.

Numerosos estudios han demostrado que la difusión de información falsa no responde únicamente a factores tecnológicos, sino también a mecanismos psicológicos y sociales que condicionan la forma en que las personas procesan la información

Desde la perspectiva del análisis del discurso, la desinformación no puede entenderse únicamente como la difusión de contenidos falsos. Con frecuencia, las estrategias de manipulación se apoyan en información parcialmente verdadera, presentada fuera de contexto, exagerada, fragmentada o construida mediante determinados recursos narrativos destinados a provocar determinadas emociones o comportamientos. En consecuencia, el estudio de la desinformación requiere analizar simultáneamente los elementos clásicos del proceso comunicativo —emisor, receptor, canal, código, mensaje y contexto—, así como los factores cognitivos y tecnológicos que intervienen en la interpretación de los mensajes.

Con frecuencia, las estrategias de manipulación se apoyan en información parcialmente verdadera, presentada fuera de contexto, exagerada, fragmentada o construida mediante determinados recursos narrativos destinados a provocar determinadas emociones o comportamientos

En este contexto, el pensamiento crítico adquiere una relevancia creciente como competencia esencial para desenvolverse en sociedades altamente digitalizadas. Más allá de la alfabetización mediática tradicional, resulta necesario desarrollar capacidades que permitan analizar la credibilidad de las fuentes, contextualizar los datos, reconocer los propios sesgos cognitivos y comprender el funcionamiento de los algoritmos que median el acceso a la información.

El presente trabajo analiza las narrativas que sustentan los procesos de desinformación en los medios digitales y se propone un modelo de análisis del discurso orientado a la verificación crítica de la información, denominado Modelo ABCD (Analizar, Buscar, Contrastar y Dudar), que ofrece un marco metodológico para identificar estrategias discursivas de manipulación y contribuir a la adquisición de competencias de alfabetización mediática adaptadas al mundo digital.

Finalmente, se plantea que la creciente automatización de la producción informativa y el desarrollo de tecnologías generativas están desplazando el foco desde la tradicional sociedad de la información hacia lo que podría denominarse una sociedad de la verificación, en la que la capacidad para evaluar críticamente la credibilidad de la información se convierte en un recurso estratégico tanto para los ciudadanos como para las instituciones democráticas.

2. Estado de la cuestión

La proliferación de la desinformación en los entornos digitales se ha consolidado como uno de los principales objetos de estudio en las ciencias de la comunicación, la psicología cognitiva, la ciencia política y la ciencia de la información durante la última década (**Benkler; Faris; Roberts, 2018; Lazer et al., 2018**). El incremento exponencial del uso de las redes sociales como principal fuente de información, unido a la disponibilidad de tecnologías de inteligencia artificial capaces de generar contenidos sintéticos de alta calidad, ha transformado profundamente los mecanismos tradicionales de la construcción de la opinión pública.

Aunque la manipulación informativa ha acompañado históricamente a las sociedades humanas, la actual infraestructura digital ha modificado sustancialmente su escala, velocidad y capacidad de influencia. La desinformación ya no depende exclusivamente de grandes organizaciones mediáticas o de campañas institucionales, sino que puede ser generada, amplificada y distribuida por millones de usuarios conectados entre sí mediante plataformas digitales gobernadas por algoritmos de recomendación (Castells, 2009; Van Dijck et al., 2018).

Uno de los principales consensos existentes en la bibliografía científica es la necesidad de diferenciar conceptualmente entre distintos tipos de información problemática. Wardle y Derakhshan (2017) proponen distinguir entre "*misinformation*", entendida como información falsa compartida sin intención de causar daño; "*disinformation*", información deliberadamente falsa creada para manipular o perjudicar; y "*malinformation*", información veraz utilizada fuera de contexto con el objetivo de causar daño o desacreditar a personas e instituciones. Esta clasificación ha permitido superar la simplificación asociada al término *fake news*, ampliamente cuestionado por su escasa precisión conceptual (Tandoc et al., 2018).

Desde una perspectiva comunicativa, diversos autores coinciden en señalar que la desinformación constituye un fenómeno narrativo antes que estrictamente tecnológico. Van Dijk (1998; 2008) sostiene que el discurso no solo transmite información, sino que contribuye activamente a la construcción de representaciones sociales, relaciones de poder e ideologías. En esta línea, Fairclough (1995) considera que los discursos constituyen prácticas sociales capaces de reproducir estructuras de dominación mediante determinadas estrategias lingüísticas y narrativas.

Las investigaciones sobre *framing* o encuadre han reforzado esta perspectiva al demostrar que la forma en que se presenta una información influye decisivamente en la interpretación realizada por la audiencia. Entman (1993) define el encuadre como el proceso mediante el cual determinados aspectos de la realidad son seleccionados y enfatizados para promover una interpretación concreta de los acontecimientos. Del mismo modo, la teoría de la *agenda-setting* desarrollada por McCombs y Shaw (1972) evidenció que los medios no determinan directamente lo que las personas piensan, pero sí influyen significativamente en los temas sobre los que piensan y debaten.

La irrupción de las plataformas digitales ha ampliado estas dinámicas mediante algoritmos capaces de personalizar la información recibida por cada usuario. Diversas investigaciones han mostrado que los sistemas de recomendación priorizan contenidos con mayor capacidad para generar interacción, favoreciendo la difusión de mensajes emocionalmente intensos frente a aquellos más rigurosos o contextualizados (Bakshy et al., 2015). Este fenómeno ha sido relacionado con la aparición de cámaras de eco (*echo chambers*) y burbujas de filtrado (*filter bubbles*), donde la exposición selectiva reduce la diversidad informativa y fortalece procesos de polarización ideológica (Sunstein, 2018).

Paralelamente, la psicología cognitiva ha aportado un marco explicativo fundamental para comprender por qué las personas aceptan y difunden contenidos falsos. Kahneman (2012) distingue entre un procesamiento rápido, intuitivo y emocional (Sistema

1) y otro más lento, analítico y deliberativo (Sistema 2). En contextos de sobrecarga informativa, el predominio del procesamiento automático favorece la utilización de sesgos cognitivos que simplifican la toma de decisiones e incrementan la vulnerabilidad a la manipulación. Entre los sesgos más estudiados destacan el sesgo de confirmación, la heurística de disponibilidad, el efecto halo, el anclaje y el efecto de verdad ilusoria, todos ellos ampliamente relacionados con la difusión de desinformación (**Pennycook; Rand**, 2019; **Lewandowsky et al.**, 2017).

Los estudios empíricos sobre propagación de noticias falsas han demostrado, además, que la viralidad responde más a factores emocionales que a la veracidad objetiva de los contenidos. El trabajo de **Vosoughi, Roy y Aral** (2018), publicado en *Science*, evidenció que las noticias falsas alcanzan una difusión significativamente mayor, más rápida y más amplia que las verdaderas en plataformas como X, un fenómeno que los autores atribuyen principalmente al comportamiento humano, especialmente cuando las informaciones generan sorpresa, indignación o miedo.

La pandemia de COVID-19 constituyó un punto de inflexión en el estudio de la desinformación. La *Organización Mundial de la Salud* acuñó el término "infodemia", como dijimos, para describir la coexistencia de información científica rigurosa con rumores, teorías conspirativas y contenidos falsos difundidos masivamente a través de redes sociales, lo que dificultó la toma de decisiones sanitarias y favoreció comportamientos de riesgo (*World Health Organization*, 2020).

La implementación de la inteligencia artificial generativa ha abierto una nueva línea de investigación centrada en los denominados *deepfakes* y en la automatización de la producción de contenidos. Algunos estudios (**Mirsky; Lee**, 2021; **Westerlund**, 2019) advierten de que la capacidad para generar imágenes, audios y vídeos hiperrealistas cuestiona la tradicional confianza depositada en la evidencia audiovisual. Paralelamente, diversos organismos internacionales, entre ellos la *UNESCO* (2023), la *Comisión Europea - European Union* (2024) y la *OECD* (2024), han subrayado la necesidad de desarrollar marcos regulatorios y programas de alfabetización digital que permitan preservar la integridad del espacio informativo.

Frente a aproximaciones centradas exclusivamente en la detección automática de noticias falsas o en el análisis de plataformas digitales, este trabajo propone un modelo interdisciplinar de análisis del discurso orientado a la verificación crítica de la información, que integra los elementos clásicos del proceso comunicativo (emisor, receptor, canal, código, mensaje y contexto) con las aportaciones de la psicología cognitiva y de la alfabetización mediática: el Modelo ABCD (Analizar, Buscar, Contrastar y Dudar).

3. Análisis de las narrativas digitales y del discurso

3.1. Las narrativas digitales como mecanismo de construcción de la realidad

Durante décadas, los medios de comunicación tradicionales actuaban como principales intermediarios en la producción y difusión de información. El ecosistema digital ha descentralizado esta función, permitiendo que cualquier usuario pueda producir, modificar y distribuir contenidos potencialmente virales a escala global (**Castells**,

2009), incrementando la pluralidad de voces pero también reduciendo los mecanismos tradicionales de control editorial y verificación.

En este contexto, la información deja de entenderse únicamente como una transmisión objetiva de hechos para convertirse en un proceso de construcción narrativa. Los acontecimientos adquieren significado mediante relatos que seleccionan determinados elementos, omiten otros y establecen relaciones causales que condicionan la interpretación realizada por la audiencia. Como señala **Van Dijk** (1998), el discurso constituye un mecanismo fundamental de construcción de la realidad social, ya que no solo comunica acontecimientos, sino que organiza cognitivamente la forma en que estos son comprendidos.

Las narrativas digitales presentan características diferenciales respecto a los modelos tradicionales de comunicación: inmediatez, hipertextualidad, interacción permanente entre emisores y receptores, fragmentación de los mensajes y personalización algorítmica de los contenidos. Estas características favorecen una circulación extremadamente rápida de la información, pero también incrementan las posibilidades de manipulación mediante descontextualización, simplificación y manipulación emocional.

La viralidad, en este sentido, no depende necesariamente de la calidad o veracidad del contenido, sino de su capacidad para captar la atención del usuario y provocar reacciones emocionales intensas. Las plataformas digitales priorizan aquellos mensajes capaces de generar interacciones —comentarios, reenvíos o reacciones— porque constituyen el principal indicador utilizado por los algoritmos para maximizar el tiempo de permanencia de los usuarios y, por tanto, la rentabilidad económica de las plataformas (**Bakshy et al.**, 2015; **Aral**, 2020).

3.2. El discurso como estrategia de influencia

Desde la perspectiva del análisis crítico del discurso, la desinformación no puede reducirse a la simple difusión de contenidos falsos. En numerosas ocasiones, las estrategias de manipulación utilizan información parcialmente verdadera presentada de manera selectiva, fuera de contexto o mediante determinados recursos lingüísticos orientados a influir en la interpretación del receptor.

Fairclough (1995) sostiene que todo discurso constituye simultáneamente una práctica textual, discursiva y social: los mensajes deben analizarse no solo por su contenido, sino también por las relaciones de poder, los intereses y los procesos sociales que condicionan su producción y circulación. En consecuencia, el análisis de una narrativa digital debe atender a cuestiones como qué aspectos de la realidad se seleccionan, qué información se omite, qué emociones pretende despertar el relato, qué interpretación se favorece y qué comportamiento se espera del receptor. En muchos casos, la eficacia de una campaña de desinformación no reside en la invención completa de los hechos, sino en la selección estratégica de fragmentos de información que permiten construir un determinado marco interpretativo (*framing*).

3.3. Arquitectura narrativa de la desinformación

Existen investigaciones (**Zannettou et al.**, 2019) que muestran que las campañas de desinformación suelen compartir patrones narrativos relativamente estables, entre los que destacan:

- simplificación extrema de problemas complejos;
- identificación de culpables claramente definidos;
- utilización de emociones intensas como miedo, indignación o ira;
- apelación constante a la urgencia;
- utilización de supuestas autoridades;
- repetición sistemática del mensaje;
- construcción de relatos dicotómicos («nosotros» frente a «ellos»).

Estos elementos facilitan el procesamiento cognitivo de la información reduciendo el esfuerzo analítico requerido para interpretar los acontecimientos. La narrativa se convierte así en un instrumento de influencia más eficaz que la simple exposición de datos objetivos: como ha demostrado la psicología cognitiva, las personas recuerdan mejor las historias que las estadísticas y procesan con mayor facilidad los relatos coherentes que la información fragmentada.

3.4. Los seis elementos del análisis discursivo

A partir del modelo clásico de la comunicación y de las aportaciones del análisis crítico del discurso, se propone analizar cualquier contenido digital mediante seis dimensiones interrelacionadas, que constituyen a su vez la base del Modelo ABCD desarrollado en el apartado 7 (figura 1).

Emisor. El análisis del emisor trasciende la simple identificación de quién publica el contenido; resulta necesario evaluar su credibilidad, intereses, financiación, posibles conflictos de interés y objetivos comunicativos. La pregunta fundamental deja de ser «¿Quién lo dice?» para convertirse en «¿Qué pretende conseguir quien lo dice?». El análisis debe incorporar tanto actores individuales como organizaciones, medios de comunicación, grupos de presión, empresas tecnológicas o redes automatizadas de difusión.

Receptor. La interpretación de un mismo mensaje puede variar considerablemente entre individuos debido a factores culturales, ideológicos y cognitivos. El receptor no constituye un sujeto pasivo, sino un agente que reconstruye el significado del mensaje a partir de sus conocimientos previos, valores, emociones y sesgos cognitivos: ¿qué emociones despierta el mensaje?, ¿confirma creencias previas?, ¿favorece respuestas impulsivas?

Canal. El medio condiciona profundamente el significado del mensaje. Una misma información publicada en una revista científica, un periódico nacional, *TikTok* o *WhatsApp* posee niveles de credibilidad y mecanismos de verificación completamente diferentes, y los algoritmos de recomendación modifican la visibilidad de los contenidos en función de criterios parcialmente opacos para los usuarios.

Código. El lenguaje constituye uno de los principales instrumentos de persuasión. Los mensajes manipulativos suelen presentar características comunes: lenguaje emocional, titulares categóricos, afirmaciones absolutas, ausencia de matices, uso estratégico de imágenes y apelaciones constantes a la indignación. El código incluye tanto elementos verbales como visuales, audiovisuales y simbólicos.

Mensaje. El análisis no debe limitarse a determinar si la información es verdadera o falsa; resulta más relevante identificar la conclusión a la que pretende conducir al receptor. Toda narrativa persigue una determinada finalidad —modificar creencias, reforzar identidades, generar miedo, aumentar la polarización, incentivar determinadas conductas—, por lo que la cuestión central pasa a ser: «¿Qué quiere que piense o haga el receptor después de consumir este contenido?».

Contexto. El contexto constituye probablemente el elemento más infravalorado en la detección de la desinformación. Una información completamente verdadera puede resultar profundamente engañosa cuando se presenta fuera de su contexto temporal, omite antecedentes relevantes, selecciona únicamente determinados datos, ignora tendencias históricas o modifica el significado original mediante un nuevo marco interpretativo. La descontextualización representa una de las estrategias más frecuentes de manipulación en los medios digitales.

3.5. De la verificación de hechos al análisis crítico del discurso

Tradicionalmente, la lucha contra la desinformación se ha centrado en los procesos de *fact-checking*, orientados a verificar la exactitud de afirmaciones concretas. Aunque estas iniciativas resultan imprescindibles, presentan limitaciones cuando la manipulación no consiste en inventar hechos, sino en reorganizar narrativamente información verdadera para inducir interpretaciones erróneas.

En este contexto, la verificación debería ampliarse hacia un análisis integral del discurso que considere simultáneamente los elementos estructurales del proceso comunicativo, los factores psicológicos del receptor y la arquitectura algorítmica de las plataformas digitales, entendiendo la desinformación no como una propiedad inherente de determinados contenidos, sino como el resultado de la interacción entre narrativas, emociones, sesgos cognitivos, incentivos económicos y dinámicas tecnológicas. Desde este enfoque, el verdadero desafío de la alfabetización mediática no consiste únicamente en enseñar a distinguir entre información verdadera y falsa, sino en desarrollar competencias que permitan interpretar críticamente los procesos mediante los cuales los discursos construyen la realidad social.

4. Papel de los sesgos cognitivos

La proliferación de la desinformación en el ecosistema digital no puede explicarse únicamente por el desarrollo tecnológico ni por la existencia de actores interesados en manipular la opinión pública. Numerosas investigaciones han demostrado que la eficacia de los procesos de desinformación depende, en gran medida, de los mecanismos cognitivos mediante los cuales las personas perciben, interpretan y recuerdan la información (Kahneman, 2012; Lewandowsky et al., 2017; Pennycook; Rand, 2019).

Desde la psicología cognitiva se reconoce que el ser humano dispone de recursos limitados para procesar la enorme cantidad de información que recibe diariamente. Para reducir esta complejidad, el cerebro utiliza heurísticos, estrategias mentales rápidas que permiten tomar decisiones con un bajo coste cognitivo. Aunque estos mecanismos resultan imprescindibles para desenvolverse en entornos complejos, también generan errores sistemáticos de juicio, conocidos como sesgos cognitivos (**Tversky; Kahneman**, 1974).

En el contexto de los medios digitales, estos sesgos adquieren una especial relevancia debido a la velocidad con la que circula la información y al diseño de plataformas que priorizan la inmediatez frente a la reflexión. La sobrecarga informativa favorece un procesamiento intuitivo basado en el denominado Sistema 1 descrito por **Kahneman** (2012), automático, rápido y emocional, frente al Sistema 2, responsable del razonamiento analítico y deliberativo, que requiere mayor esfuerzo cognitivo y suele activarse con menor frecuencia en entornos digitales altamente acelerados. Esto explica en parte por qué la desinformación logra difundirse con tanta eficacia: como demostraron **Vosoughi, Roy y Aral** (2018), las noticias falsas se propagan más rápidamente que las verdaderas no debido a una mayor actividad de los bots, sino porque despiertan emociones intensas —sorpresa, miedo o indignación— que incrementan la probabilidad de ser compartidas por los propios usuarios.

4.1. Sesgo de confirmación: la búsqueda de información coherente

Entre los sesgos más estudiados destaca el sesgo de confirmación, definido como la tendencia a buscar, interpretar y recordar preferentemente aquella información que confirma nuestras creencias previas, ignorando o minimizando la información contradictoria (**Nickerson**, 1998). En el ecosistema digital este sesgo se ve amplificado por los algoritmos de recomendación, que personalizan los contenidos en función del historial de navegación, las interacciones y las preferencias del usuario, favoreciendo la formación de cámaras de eco (*echo chambers*) y reduciendo la diversidad informativa disponible.

4.2. Heurística de disponibilidad y la percepción del riesgo

Otro mecanismo especialmente relevante es la heurística de disponibilidad, mediante la cual las personas estiman la probabilidad de un acontecimiento en función de la facilidad con la que recuerdan ejemplos similares (**Tversky; Kahneman**, 1974). Las narrativas digitales explotan este mecanismo mediante la repetición constante de sucesos especialmente llamativos o emocionalmente impactantes: un hecho aislado puede llegar a percibirse como un fenómeno generalizado simplemente porque aparece reiteradamente en redes sociales o medios de comunicación, distorsionando la percepción social del riesgo respecto a fenómenos como la inmigración, la delincuencia, las pandemias o determinados conflictos políticos.

4.3. Efecto halo y la credibilidad de las fuentes

El efecto halo describe la tendencia a extrapolar una característica positiva de una persona o institución al conjunto de sus actuaciones (**Thorndike**, 1920). En el ámbito digital este fenómeno explica por qué numerosos usuarios otorgan credibilidad a contenidos difundidos por personas famosas, influencers o líderes de opinión, independientemente de que posean conocimientos especializados sobre la materia tratada:

la popularidad, el prestigio profesional o el número de seguidores pueden actuar como indicadores heurísticos de credibilidad, sustituyendo la evaluación crítica de las evidencias presentadas.

4.4. Anclaje y la persistencia de la primera impresión

El sesgo de anclaje hace referencia a la influencia desproporcionada que ejerce la primera información recibida sobre las valoraciones posteriores (**Tversky; Kahneman, 1974**). En la práctica, las primeras versiones de un acontecimiento suelen condicionar la interpretación de las informaciones posteriores, incluso cuando estas corrigen errores iniciales, lo que explica parcialmente por qué los procesos de verificación resultan menos eficaces que la difusión inicial del bulo.

4.5. Efecto de verdad ilusoria y repetición narrativa

La repetición constituye uno de los mecanismos más eficaces para incrementar la credibilidad percibida de una afirmación. El denominado efecto de verdad ilusoria describe la tendencia a considerar más verdadera una información simplemente por haber sido expuesta repetidamente, incluso cuando se sabe que carece de fundamento. Las plataformas digitales potencian este fenómeno mediante algoritmos que favorecen la circulación reiterada de contenidos con elevada interacción, incrementando la familiaridad del mensaje y, con ello, su apariencia de veracidad. Desde una perspectiva comunicativa, este efecto explica por qué muchas campañas de desinformación no necesitan aportar nuevas evidencias, sino únicamente mantener una presencia constante en el espacio público.

4.6. Emociones y toma de decisiones

Se ha demostrado que las emociones desempeñan un papel central en la difusión de contenidos digitales (**Bakir; McStay, 2018**). El miedo, la ira, la sorpresa o la indignación reducen la capacidad de razonamiento analítico y favorecen respuestas impulsivas orientadas a compartir información antes de verificarla. Las narrativas manipulativas utilizan frecuentemente titulares alarmistas, lenguaje polarizador y apelaciones a la urgencia precisamente porque estos recursos incrementan la activación emocional y reducen el procesamiento crítico del mensaje: la viralidad depende menos de la exactitud de la información que de su capacidad para movilizar respuestas afectivas.

4.7. Sesgos cognitivos, alfabetización mediática y pensamiento crítico

La existencia de sesgos cognitivos no implica que las personas sean irracionales ni incapaces de identificar información falsa: son adaptaciones evolutivas que permiten procesar grandes volúmenes de información con rapidez y eficiencia. Sin embargo, en un entorno caracterizado por la sobreabundancia informativa y la personalización algorítmica, estos mismos mecanismos pueden convertirse en factores de vulnerabilidad frente a la manipulación. Por ello, la alfabetización me-

La pregunta crítica deja de ser únicamente «¿Es cierta esta información?» para incorporar una cuestión previa igualmente relevante: «¿Por qué estoy dispuesto a creerla?». Este cambio de enfoque implica desacelerar el procesamiento automático de la información y favorecer estrategias de análisis consciente

diática contemporánea no debería limitarse a enseñar procedimientos de verificación documental o identificación de fuentes fiables, sino también desarrollar competencias metacognitivas que permitan reconocer cómo influyen los propios sesgos en la interpretación de la información. En este sentido, la pregunta crítica deja de ser únicamente «¿Es cierta esta información?» para incorporar una cuestión previa igualmente relevante: «¿Por qué estoy dispuesto a creerla?». Este cambio de enfoque implica desacelerar el procesamiento automático de la información y favorecer estrategias de análisis consciente que permitan identificar emociones, cuestionar interpretaciones iniciales y contrastar las evidencias disponibles: la alfabetización digital debe entenderse, ante todo, como una competencia cognitiva antes que tecnológica.

5. IA generativa y deepfakes: nuevos desafíos para la verificación de la información

La inteligencia artificial (IA) generativa es uno de los avances tecnológicos más relevantes de la última década. Basada en los llamados modelos de aprendizaje profundo (*deep learning*) entrenados con grandes volúmenes de datos, esta tecnología permite generar textos, imágenes, audios, vídeos e incluso código informático con un grado de calidad que, en muchos casos, resulta difícil de distinguir de la producción humana (Bommasani et al., 2021; Kaplan et al., 2020). Su rápida adopción está transformando sectores tan diversos como la educación, la medicina, la investigación científica, la industria creativa y la comunicación.

Sin embargo, el impacto de la IA generativa trasciende el ámbito tecnológico: su capacidad para producir contenidos sintéticos altamente verosímiles plantea nuevos interrogantes sobre la autenticidad de la información, la confianza en las evidencias digitales y la integridad del debate público. Como toda tecnología de propósito general, la IA amplifica tanto las oportunidades como los riesgos: facilita procesos de innovación, pero también puede ser utilizada para automatizar estrategias de manipulación informativa, suplantación de identidad y generación de contenidos engañosos.

5.1. De la automatización de tareas a la automatización de la realidad

Las primeras aplicaciones de la inteligencia artificial se orientaban principalmente hacia tareas de clasificación, predicción o reconocimiento de patrones. Los sistemas actuales incorporan una capacidad adicional: la generación autónoma de contenidos. Modelos de lenguaje, generadores de imágenes y sintetizadores de voz permiten crear materiales originales a partir de instrucciones en lenguaje natural, reduciendo drásticamente las barreras técnicas para producir contenidos digitales.

Esta evolución supone un cambio cualitativo en el ecosistema comunicativo. Si anteriormente la manipulación audiovisual requería conocimientos especializados y elevados recursos técnicos, hoy cualquier usuario puede generar imágenes hiperrealistas o vídeos sintéticos utilizando herramientas accesibles y de bajo coste. En consecuencia, el principal desafío ya no reside únicamente en distinguir entre contenidos profesionales y *amateur*, sino en determinar si un contenido corresponde a un acontecimiento real o ha sido generado artificialmente.

5.2. Deepfakes y crisis de la evidencia audiovisual

Entre las aplicaciones más controvertidas de la IA generativa destacan los *deepfakes*, entendidos como imágenes, vídeos o audios sintéticos creados mediante redes neuronales capaces de reproducir con gran precisión la apariencia, la voz o los gestos de una persona (**Mirsky; Lee**, 2021). Aunque las primeras investigaciones sobre *deepfakes* se centraron en aplicaciones recreativas o experimentales, su utilización se ha extendido rápidamente hacia ámbitos como la desinformación política, el fraude financiero, la pornografía no consentida, la suplantación de identidad y las campañas de manipulación electoral. La creciente calidad de estos contenidos dificulta su detección incluso por observadores entrenados, cuestionando uno de los principios tradicionales de la comunicación: la confianza en la evidencia visual.

Durante décadas, fotografías y grabaciones audiovisuales fueron consideradas pruebas privilegiadas de la realidad. La IA generativa altera profundamente este paradigma: la capacidad para fabricar escenas completamente ficticias con apariencia documental convierte la imagen en un elemento que también requiere verificación. La relación entre imagen y verdad deja de ser inmediata, y la credibilidad ya no puede sustentarse exclusivamente en la apariencia visual, sino que exige contextualizar el origen, la trazabilidad y la autenticidad del contenido.

5.3. Dividendo del mentiroso y la erosión de la confianza

Uno de los efectos más preocupantes de la expansión de los *deepfakes* es el denominado dividendo del mentiroso (*liar's dividend*), concepto que describe la posibilidad de desacreditar pruebas auténticas alegando que han sido generadas mediante inteligencia artificial (**Chesney; Citron**, 2019). Este fenómeno introduce una paradoja relevante: cuanto más sofisticadas son las herramientas para fabricar contenidos falsos, más sencillo resulta cuestionar la autenticidad de los contenidos verdaderos, ya que, a priori, las diferencias entre información falsa y verdadera no son tan evidentes. La consecuencia no es únicamente un aumento de la desinformación, sino una erosión progresiva de la confianza social en cualquier evidencia digital, que afecta tanto a los medios de comunicación como a las instituciones públicas, los sistemas judiciales y las relaciones interpersonales.

5.4. Inteligencia artificial, algoritmos y amplificación de la desinformación

La IA no solo interviene en la creación de contenidos, sino también en su distribución. Las principales plataformas digitales utilizan algoritmos de recomendación basados en aprendizaje automático para seleccionar la información que recibe cada usuario, optimizando variables como el tiempo de permanencia, la interacción o la probabilidad de compartir contenidos. Diversos estudios, como el de **Zhou y Zafarani** (2020), han señalado que estos algoritmos tienden a favorecer contenidos emocionalmente intensos, sorprendentes o controvertidos, ya que generan mayores niveles de interacción; sin que exista necesariamente una intención explícita de promover la desinformación, el diseño de estos sistemas puede contribuir indirectamente a amplificar narrativas polarizadoras o engañosas.

5.5. Estrategias para preservar la integridad del ecosistema informativo

La respuesta frente a los riesgos asociados a la IA generativa requiere una aproximación multidimensional que combine soluciones tecnológicas, regulatorias y educativas. En el ámbito tecnológico se están implementando programas de detección automática de contenidos sintéticos, sistemas de autenticación criptográfica y estándares de procedencia digital, como las iniciativas promovidas por la *Coalition for Content Provenance and Authenticity (C2PA)*, orientadas a certificar el origen y las modificaciones realizadas sobre imágenes y vídeos. No obstante, estas soluciones se enfrentan a una carrera constante con la mejora de los modelos generativos, lo que limita su eficacia a largo plazo.

Desde el punto de vista normativo, la *Ley Europea de Inteligencia Artificial (AI Act)* constituye el primer marco regulador integral destinado a establecer obligaciones de transparencia para determinados sistemas de IA, incluyendo requisitos específicos para la identificación de contenidos generados artificialmente y la gestión de riesgos asociados a aplicaciones de alto impacto (*European Union, 2024*). Sin embargo, la bibliografía científica coincide en que ninguna solución exclusivamente tecnológica o jurídica resultará suficiente: la resiliencia frente a la desinformación depende, en última instancia, de la capacidad de la ciudadanía para interpretar críticamente la información, comprender el funcionamiento de los sistemas algorítmicos y desarrollar competencias de alfabetización mediática adaptadas a la era de la IA.

5.6. De la sociedad de la información a la sociedad de la verificación

La expansión de la inteligencia artificial generativa obliga a replantear algunos de los principios tradicionales sobre los que se ha construido el ecosistema informativo. Durante décadas, el principal desafío consistía en garantizar el acceso a la información; posteriormente, la atención se desplazó hacia la gestión del exceso de datos y la denominada infoxicación. En la actualidad, el reto fundamental consiste en determinar la autenticidad y fiabilidad de los contenidos que consumimos, lo que permite plantear que las sociedades contemporáneas evolucionan desde una sociedad de la información hacia una sociedad de la verificación, en la que el valor estratégico ya no reside únicamente en disponer de información, sino en contar con mecanismos eficaces para evaluar su credibilidad. En este nuevo contexto, la verificación se convierte en una competencia transversal que implica la participación activa de instituciones, plataformas, sistemas educativos y ciudadanía.

El reto fundamental consiste en determinar la autenticidad y fiabilidad de los contenidos que consumimos, lo que permite plantear que las sociedades contemporáneas evolucionan desde una sociedad de la información hacia una sociedad de la verificación, en la que el valor estratégico ya no reside únicamente en disponer de información, sino en contar con mecanismos eficaces para evaluar su credibilidad

6. Casos de estudio

La complejidad del fenómeno de la desinformación exige analizar casos concretos que permitan observar la interacción entre los elementos del discurso, los sesgos cognitivos y las dinámicas propias de las plataformas digitales. Los casos seleccionados representan diferentes modalidades de manipulación informativa y muestran cómo la desinformación constituye un fenómeno multidimensional en el que intervienen factores tecnológicos, psicológicos, económicos y políticos.

6.1. *Cambridge Analytica*: la personalización de la persuasión política

Uno de los casos más relevantes en la historia reciente de la comunicación digital fue el escándalo protagonizado por *Cambridge Analytica* durante las elecciones presidenciales de Estados Unidos de 2016 y el referéndum del Brexit. La empresa obtuvo datos personales de millones de usuarios de *Facebook* sin su consentimiento para elaborar perfiles psicográficos y diseñar campañas de comunicación altamente segmentadas. Más allá del uso indebido de datos personales, este caso evidenció que la influencia política podía ejercerse mediante la personalización extrema de los mensajes: en lugar de difundir un único discurso dirigido a toda la población, se produjeron múltiples narrativas adaptadas a las características psicológicas de cada segmento de usuarios, explotando sesgos como el de confirmación o la aversión a la pérdida.

6.2. La COVID-19 y la infodemia global

La pandemia de COVID-19 supuso uno de los mayores desafíos informativos de la historia reciente. Paralelamente a la expansión del virus SARS-CoV-2, se produjo una difusión masiva de rumores, teorías conspirativas y falsas recomendaciones sanitarias que la *Organización Mundial de la Salud* denominó infodemia. Las investigaciones llevadas a cabo durante este periodo documentaron miles de informaciones falsas relacionadas con el origen del virus, tratamientos sin evidencia científica, eficacia de las vacunas o supuestas conspiraciones internacionales, contenidos que compartían características comunes: apelaciones constantes al miedo y a la incertidumbre, utilización de testimonios personales como pretendida evidencia, referencias a falsas autoridades científicas e interpretación conspirativa de acontecimientos complejos. La pandemia puso de manifiesto que la alfabetización científica es un elemento inseparable de la alfabetización mediática.

6.3. Los disturbios en el Reino Unido: cuando la desinformación salta de la pantalla a la calle

Un ejemplo del impacto social de las narrativas digitales se produjo tras el asesinato de tres niñas en Southport (Reino Unido) en 2024. Poco después del ataque comenzaron a difundirse en redes sociales numerosos mensajes que atribuían falsamente la autoría a un inmigrante musulmán, acompañados de nombres inventados y fotografías manipuladas. Aunque la información fue desmentida por las autoridades en pocas horas, la velocidad de propagación de los rumores favoreció la organización de movilizaciones violentas y ataques contra comunidades inmigrantes en varias ciudades británicas. El análisis del discurso muestra cómo la información inicial fue reinterpretada mediante un marco narrativo previamente existente sobre inmigración, seguridad y conflicto cultural, reforzando prejuicios y activando mecanismos emocionales de indignación colectiva a una velocidad que superó ampliamente la capacidad de respuesta de los procesos tradicionales de verificación.

6.4. El fenómeno *Pizzagate*: la construcción narrativa de una teoría conspirativa

Entre los ejemplos más representativos de las teorías conspirativas difundidas en redes sociales destaca el denominado *Pizzagate*. Durante la campaña presidencial estadounidense de 2016 comenzaron a circular mensajes que afirmaban, sin ninguna evidencia, que una conocida pizzería de Washington D. C. albergaba una red internacional de explotación infantil que se vinculaba a Hillary Clinton. La difusión reiterada de esta narrativa culminó cuando un ciudadano armado accedió al establecimiento convencido de que iba a liberar a las supuestas víctimas. El caso evidencia cómo la repetición constante de una narrativa coherente puede generar una percepción subjetiva de veracidad incluso en ausencia absoluta de pruebas, y constituye un ejemplo paradigmático del efecto de verdad ilusoria, del sesgo de confirmación y del papel de las comunidades digitales en la construcción de realidades compartidas al margen de la evidencia.

6.5. *Deepfakes* y manipulación audiovisual: el caso del presidente Zelenski

La expansión de la inteligencia artificial generativa ha incorporado una nueva dimensión al fenómeno de la desinformación. Durante la invasión rusa de Ucrania comenzó a circular un vídeo manipulado mediante técnicas de *deepfake* en el que el presidente Volodímir Zelenski aparentemente anunciaba la rendición del ejército ucraniano. Aunque la baja calidad técnica permitió detectar rápidamente la falsificación, el caso anticipó un escenario en el que la manipulación audiovisual puede alcanzar niveles prácticamente indistinguibles de la realidad. El interés de este ejemplo no reside únicamente en el contenido falso, sino en la alteración de uno de los principales mecanismos de confianza social —la evidencia audiovisual—: cuando cualquier imagen puede ser generada artificialmente, la verificación deja de depender de la observación directa y requiere analizar el origen, la trazabilidad y el contexto de producción del contenido.

6.6. Análisis comparado de los casos

Los cinco casos analizados presentan diferencias en cuanto a contexto histórico, tecnología utilizada y objetivos perseguidos, pero comparten un conjunto de elementos estructurales que permiten identificar patrones comunes en las narrativas digitales de desinformación:

- En primer lugar, ninguno se basa exclusivamente en la invención completa de hechos: en todos ellos existe un acontecimiento real —unas elecciones, una pandemia, un atentado, un establecimiento físico o un conflicto bélico— sobre el que posteriormente se construye una narrativa manipulada mediante la selección, reinterpretación o descontextualización de determinados elementos.
- En segundo lugar, la eficacia de estas narrativas depende menos de su veracidad objetiva que de su capacidad para activar emociones intensas, reforzar identidades colectivas y confirmar creencias previamente existentes, de modo que los sesgos cognitivos del receptor desempeñan un papel tan relevante como la propia estructura del mensaje.
- Finalmente, todos los casos ponen de manifiesto que la tecnología constituye un elemento facilitador, pero no la causa principal del fenómeno: la desinfor-

mación surge de la interacción entre intereses estratégicos, arquitecturas algorítmicas, vulnerabilidades cognitivas y dinámicas sociales de circulación de la información.

Desde esta perspectiva, los casos analizados refuerzan la necesidad de superar aproximaciones centradas exclusivamente en la verificación factual para avanzar hacia modelos de análisis capaces de integrar simultáneamente el estudio del discurso, los procesos cognitivos y las características del ecosistema digital. Sobre esta premisa se fundamenta la propuesta metodológica presentada en el siguiente apartado: el Modelo ABCD (Analizar, Buscar, Contrastar y Dudar).

7. Propuesta de un modelo de análisis del discurso para la verificación (ABCD)

7.1. Fundamentación del modelo

El análisis realizado en los apartados anteriores pone de manifiesto que la desinformación constituye un fenómeno complejo que no puede abordarse únicamente mediante procedimientos tradicionales de verificación factual (*fact-checking*). La creciente sofisticación de las narrativas digitales, la personalización algorítmica de los contenidos y la expansión de la inteligencia artificial generativa hacen necesario desarrollar herramientas de análisis capaces de integrar dimensiones comunicativas, cognitivas y tecnológicas.

Los modelos de verificación existentes se centran principalmente en determinar la exactitud de afirmaciones concretas, contrastando datos con fuentes documentales. Aunque esta aproximación resulta imprescindible, presenta limitaciones cuando la manipulación no consiste en inventar hechos, sino en reorganizar información veraz mediante procesos de selección, omisión, exageración o descontextualización. En este contexto se propone el Modelo ABCD (Analizar, Buscar, Contrastar y Dudar), concebido como un procedimiento metodológico orientado a fortalecer el pensamiento crítico mediante el análisis integral del discurso digital. El modelo combina los elementos clásicos del proceso comunicativo con la evaluación de los sesgos cognitivos del receptor y las características del ecosistema digital: su objetivo no consiste únicamente en determinar si una información es verdadera o falsa, sino en comprender cómo se construye el significado de los mensajes, qué intereses pueden motivar su difusión y de qué manera influyen los procesos psicológicos y tecnológicos en su interpretación.

7.2. Estructura general del Modelo

El Modelo ABCD se articula en cuatro fases secuenciales e interdependientes —Analizar, Buscar, Contrastar y Dudar— que se aplican sobre los seis elementos fundamentales del proceso comunicativo (emisor, receptor, canal, código, mensaje y contexto), configurando un procedimiento sistemático para evaluar la fiabilidad de cualquier contenido digital (figura 1).

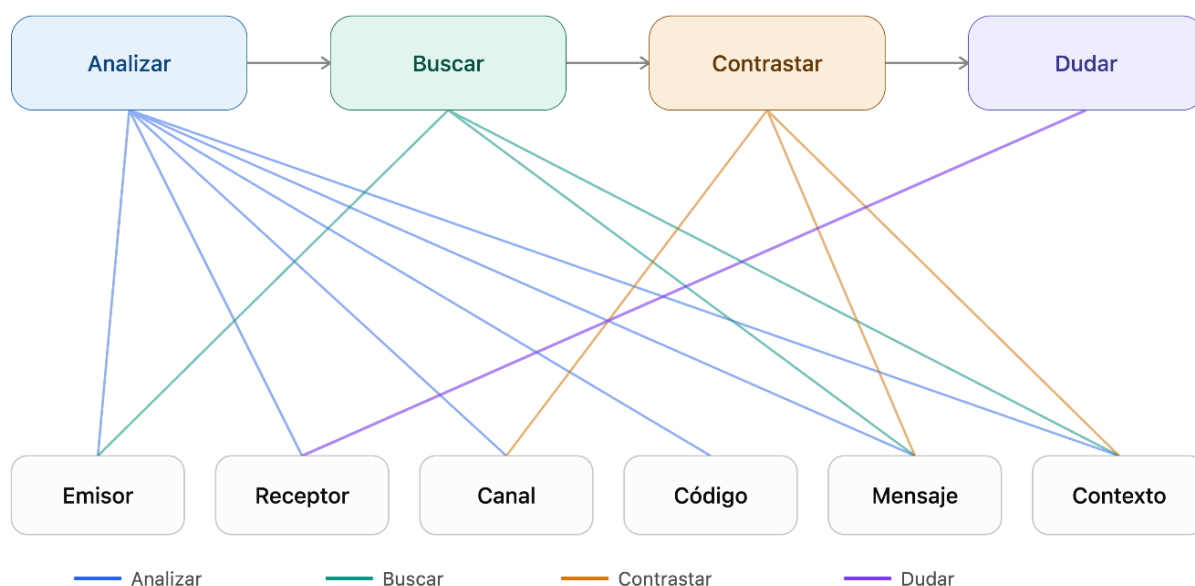


Figura 1. Estructura conceptual del Modelo ABCD.

7.3. Primera fase: Analizar

La primera fase tiene como finalidad comprender la arquitectura del mensaje antes de aceptar o rechazar su contenido, sustituyendo la reacción inmediata por una observación sistemática del discurso a través de los seis elementos comunicativos:

- Emisor: ¿Quién publica la información? ¿Es identificable? ¿Qué intereses económicos, políticos o ideológicos puede tener? ¿Obtiene algún beneficio con la difusión del mensaje?
- Receptor: ¿Qué emociones me provoca? ¿Confirma mis creencias? ¿Estoy reaccionando antes de pensar?
- Canal: ¿Dónde aparece publicada? ¿Existe supervisión editorial? ¿Qué papel desempeñan los algoritmos en su difusión?
- Código: ¿Utiliza lenguaje alarmista? ¿Exagera? ¿Simplifica un problema complejo? ¿Presenta imágenes impactantes?
- Mensaje: ¿Cuál es realmente la idea principal? ¿Qué quiere que piense? ¿Qué quiere que haga?
- Contexto: ¿Faltan datos? ¿Existe información previa? ¿Las cifras están completas? ¿Puede estar descontextualizada?

7.4. Segunda fase: Buscar

Una vez identificado el funcionamiento interno del discurso, el siguiente paso consiste en localizar información independiente. Buscar no significa únicamente utilizar un buscador, sino identificar evidencia fiable: localizar la fuente original, consultar organismos oficiales, revisar bibliografía científica, comprobar fechas, verificar imágenes mediante búsqueda inversa, analizar metadatos cuando sea posible e identificar quién financia la información. Esta fase pretende reducir la dependencia de una única fuente de información.

7.5. Tercera fase: Contrastar

El contraste es el núcleo de la verificación: no basta con encontrar una fuente alternativa, sino que es necesario comparar distintas perspectivas —consultar medios con líneas editoriales diferentes, comparar datos estadísticos, verificar documentos originales, analizar informes científicos e identificar posibles discrepancias—. Especial importancia adquiere el contexto, ya que muchas informaciones falsas utilizan datos reales presentados de forma parcial; en consecuencia, el contraste debe incorporar tanto la comprobación factual como el análisis del marco narrativo en el que esos datos aparecen integrados.

7.6. Cuarta fase: Dudar

La última fase constituye el elemento diferenciador del modelo. Dudar no significa desconfiar sistemáticamente de toda información, sino reconocer las limitaciones del propio razonamiento: en esta etapa el análisis se dirige hacia el propio receptor, con preguntas como ¿por qué quiero creer esta información?, ¿qué emociones está despertando?, ¿estoy ignorando datos contradictorios?, ¿estoy compartiendo antes de comprender?, ¿cómo influyen mis sesgos cognitivos? Esta dimensión metacognitiva diferencia el Modelo ABCD de los procedimientos tradicionales de *fact-checking*, ya que el objeto de análisis deja de ser únicamente el mensaje para incorporar al propio sujeto que interpreta la información.

7.7. Integración del modelo con el análisis del discurso

El Modelo ABCD se fundamenta sobre los seis elementos clásicos del proceso comunicativo, cada uno de los cuales incorpora una pregunta crítica de verificación (tabla 1):

Tabla 1. Seis elementos clásicos del proceso comunicativo

Elemento	Pregunta crítica
Emisor	¿Quién gana con este mensaje?
Receptor	¿Cómo estoy interpretándolo?
Canal	¿Por qué aparece precisamente aquí?
Código	¿Qué recursos persuasivos utiliza?
Mensaje	¿Qué quiere que piense o haga?
Contexto	¿Qué información falta?

Esta estructura permite analizar simultáneamente dimensiones comunicativas, psicológicas y tecnológicas.

7.8. Aplicación práctica y aportaciones del Modelo ABCD

El modelo puede aplicarse en distintos ámbitos: alfabetización mediática, educación secundaria y universitaria, formación de periodistas, comunicación institucional, verificación profesional, organizaciones públicas, análisis de campañas electorales e investigación académica. Su flexibilidad permite adaptarlo tanto al análisis manual como al diseño de herramientas digitales de apoyo a la verificación.

Frente a otros procedimientos centrados exclusivamente en verificar hechos, el Modelo ABCD integra el análisis del discurso y la psicología cognitiva, incorpora el papel

de los algoritmos, analiza la construcción narrativa de la información, introduce la metacognición como instrumento de verificación, puede utilizarse con información verdadera, falsa o parcialmente verdadera, y resulta aplicable en contextos educativos y profesionales. Más que identificar noticias falsas, el modelo pretende fortalecer la capacidad crítica de los ciudadanos para desenvolverse en un ecosistema comunicativo caracterizado por la sobreabundancia informativa, la automatización de los contenidos y la creciente dificultad para distinguir entre realidad y simulación.

7.9. Hacia una alfabetización para la sociedad de la verificación

La expansión de la inteligencia artificial y la progresiva automatización de la producción de contenidos hacen prever que la verificación será una competencia esencial en las próximas décadas. En este escenario, la alfabetización mediática no puede limitarse a enseñar a utilizar herramientas digitales o reconocer fuentes fiables; resulta igualmente necesario formar a la ciudadanía en metodologías para analizar críticamente los discursos, identificar los propios sesgos cognitivos y comprender el funcionamiento de los sistemas algorítmicos. El Modelo ABCD se plantea como una contribución en esta dirección: más que ofrecer un procedimiento cerrado, propone un marco de análisis adaptable a distintos contextos comunicativos y orientado a promover una ciudadanía capaz de tomar decisiones informadas en un entorno caracterizado por la incertidumbre y la sobreabundancia de información.

La alfabetización mediática no puede limitarse a enseñar a utilizar herramientas digitales o reconocer fuentes fiables; resulta igualmente necesario formar a la ciudadanía en metodologías para analizar críticamente los discursos, identificar los propios sesgos cognitivos y comprender el funcionamiento de los sistemas algorítmicos

8. Discusión

El análisis realizado en este trabajo permite afirmar que la desinformación contemporánea constituye un fenómeno complejo que no puede entenderse únicamente como la difusión de contenidos falsos. La manipulación informativa rara vez se basa exclusivamente en la invención completa de hechos; con frecuencia opera mediante estrategias más complejas como la selección parcial de datos, la descontextualización, la exageración, la asociación emocional o la reinterpretación interesada de acontecimientos reales. Por ello, el reto actual no consiste únicamente en identificar falsedades, sino en comprender cómo se construyen los significados y qué mecanismos favorecen que determinadas narrativas sean aceptadas, compartidas y amplificadas.

8.1. De la verificación factual al análisis integral del discurso

Los resultados del análisis evidencian la necesidad de ampliar el concepto tradicional de verificación. El *fact-checking* continúa siendo imprescindible, especialmente en áreas como la comunicación política, la salud pública o la información científica, pero presenta limitaciones cuando la desinformación no reside en una afirmación completamente falsa, sino en la forma en que se organiza y presenta la información: una noticia puede contener datos verdaderos y, al mismo tiempo, construir una interpretación engañosa mediante la omisión de contexto, la selección interesada de evidencias

o el uso de marcos narrativos orientados a provocar determinadas respuestas emocionales. El Modelo ABCD propuesto responde precisamente a esta necesidad al desplazar el foco desde la identificación exclusiva del error hacia la comprensión del proceso completo de construcción y recepción del mensaje.

8.2. La paradoja tecnológica: más información, menor certeza

Uno de los principales hallazgos del análisis es la existencia de una paradoja característica de la sociedad digital: nunca hemos tenido mayor acceso a información y, al mismo tiempo, nunca ha sido tan difícil establecer qué información merece confianza. Los *deepfakes*, los textos generados automáticamente y las campañas de desinformación asistidas mediante inteligencia artificial introducen la posibilidad de fabricar evidencias aparentemente auténticas. Sin embargo, el problema no reside exclusivamente en la capacidad técnica de generar falsificaciones: la amenaza más profunda aparece cuando la proliferación de contenidos sintéticos produce una pérdida generalizada de confianza, de modo que tanto lo falso como lo verdadero pueden llegar a percibirse como potencialmente manipulables (*liar's dividend*).

8.3. La dimensión cognitiva: el ciudadano como parte del ecosistema informativo

El análisis de los sesgos cognitivos muestra que la desinformación no puede explicarse únicamente por la acción de emisores malintencionados o por las características técnicas de las plataformas digitales: el receptor desempeña un papel fundamental en la circulación de las narrativas digitales, ya que los sesgos de confirmación, disponibilidad, anclaje, autoridad o verdad ilusoria condicionan tanto la interpretación de la información como la disposición a compartirla. Esto implica que la lucha contra la desinformación no puede limitarse a producir más información correcta: en un entorno saturado de datos, la cuestión fundamental es desarrollar capacidades para evaluar la información recibida, reconocer las propias limitaciones cognitivas y comprender cómo nuestras emociones influyen en nuestras decisiones.

8.4. Algoritmos, economía de la atención y polarización

Los algoritmos de recomendación no fueron diseñados inicialmente para difundir desinformación, sino para optimizar la experiencia del usuario y maximizar la interacción. Sin embargo, cuando la atención se convierte en el principal recurso económico, los contenidos que generan emociones intensas adquieren una ventaja competitiva, generando un desplazamiento desde una comunicación basada en la argumentación hacia una comunicación basada en la reacción. Como consecuencia, el espacio público digital puede favorecer procesos de polarización, reducción de la diversidad informativa y consolidación de comunidades cerradas de interpretación; el desafío no consiste únicamente en exigir responsabilidad individual al usuario, sino también en analizar críticamente los modelos económicos y tecnológicos que estructuran la circulación de la información.

8.5. Limitaciones y futuras líneas de investigación

Aunque el Modelo ABCD proporciona un marco integrador para el análisis de narrativas digitales, presenta algunas limitaciones. En primer lugar, su aplicación requiere un componente interpretativo humano, especialmente en aquellos casos donde la manipulación no depende de datos falsos, sino de marcos narrativos o estrategias discursivas más sutiles. En segundo lugar, la velocidad de evolución de la inteligencia

artificial plantea la necesidad de actualizar continuamente las metodologías de análisis, ya que las técnicas actuales de generación sintética evolucionan más rápidamente que muchos sistemas educativos y regulatorios. Finalmente, sería necesario validar empíricamente el modelo mediante estudios experimentales que permitan analizar su eficacia en diferentes grupos de población, especialmente entre jóvenes y adolescentes. Las futuras investigaciones podrían explorar la integración del Modelo ABCD con sistemas de inteligencia artificial explicable (*Explainable AI*) capaces de asistir en procesos de verificación manteniendo la supervisión humana.

9. Conclusiones

La transformación del ecosistema informativo digital ha modificado profundamente la forma en la que producimos, distribuimos e interpretamos la información. La sociedad contemporánea se caracteriza por una disponibilidad prácticamente ilimitada de datos, contenidos y discursos, pero esta abundancia no se ha traducido necesariamente en una mayor comprensión de la realidad: la saturación informativa, la aceleración de los procesos comunicativos y la personalización algorítmica han generado un nuevo desafío, el de distinguir entre información, desinformación y conocimiento.

El análisis llevado a cabo en este trabajo permite concluir que la desinformación no debe entenderse únicamente como la difusión de contenidos falsos, sino como un fenómeno mucho más complejo en el que intervienen la construcción narrativa de los mensajes, los intereses de los emisores, las características de los canales digitales, los mecanismos cognitivos de los receptores y las posibilidades tecnológicas de producción y amplificación de contenidos. Los casos analizados muestran que los bulos y las noticias falsas alcanzan capacidad de impacto social no solamente porque sean falsos, sino porque conectan con estructuras narrativas previamente existentes, activan emociones intensas y aprovechan vulnerabilidades cognitivas presentes en todos los individuos.

La desinformación no debe entenderse únicamente como la difusión de contenidos falsos, sino como un fenómeno mucho más complejo en el que intervienen la construcción narrativa de los mensajes, los intereses de los emisores, las características de los canales digitales, los mecanismos cognitivos de los receptores y las posibilidades tecnológicas de producción y amplificación de contenidos

La expansión de la inteligencia artificial generativa introduce una nueva dimensión en este escenario. La aparición de imágenes, vídeos y audios sintéticos altamente realistas cuestiona la tradicional relación entre evidencia visual y verdad. Sin embargo, la IA no constituye en sí misma el problema, sino una herramienta cuyo impacto depende del uso que se haga de ella: la misma tecnología que permite avances extraordinarios en ámbitos científicos, educativos, sanitarios o empresariales puede ser utilizada para automatizar procesos de manipulación informativa. En este contexto, la principal amenaza no reside únicamente en la existencia de contenidos falsos, sino en la progresiva erosión de la confianza social, ya que la posibilidad de fabricar evidencias falsas puede generar una situación paradójica en la que incluso los contenidos verdaderos sean cuestionados.

Como contribución principal, este artículo propone el Modelo ABCD (Analizar, Buscar, Contrastar y Dudar) como un marco metodológico para abordar la verificación desde una perspectiva integral, que incorpora el análisis del proceso comunicativo completo: quién produce el mensaje, cómo se construye, dónde circula, qué recursos persuasivos utiliza, qué emociones activa y cómo intervienen los propios sesgos del receptor. Su principal aportación reside en desplazar la pregunta tradicional de «¿es verdadero o falso este contenido?» hacia una cuestión más amplia: «¿cómo se ha construido este mensaje, qué pretende provocar y cómo está condicionando mi interpretación de la realidad?».

Asimismo, el trabajo pone de manifiesto que la alfabetización mediática debe evolucionar hacia una alfabetización crítica adaptada a la inteligencia artificial: en la actualidad no basta con saber acceder a información o utilizar herramientas digitales, sino que es necesario comprender cómo funcionan los algoritmos, cómo se construyen las narrativas digitales, cómo influyen nuestras emociones y cómo evaluar la autenticidad de los contenidos generados artificialmente. La sociedad del futuro no será aquella que disponga de mayor cantidad de información, sino la que desarrolle mejores capacidades para interpretarla; en un contexto donde la inteligencia artificial aumentará progresivamente la capacidad de producir contenidos sintéticos, el valor diferencial del ser humano residirá en competencias difícilmente automatizables: el pensamiento crítico, la contextualización, la ética, la empatía y la capacidad de formular buenas preguntas.

La transición desde la sociedad de la información hacia la sociedad de la verificación exige ciudadanos capaces de analizar, buscar, contrastar y dudar. Porque el mayor riesgo de la era digital no es vivir sin información, sino tomar decisiones individuales y colectivas basadas en información incompleta, manipulada o interpretada de manera incorrecta.

10. Referencias

Aral, S. (2020). *The Hype Machine*. Currency. ISBN: 978 0 525574583

Bakir, V.; McStay, A. (2018). Fake news and the economy of emotions. *Digital Journalism*, v. 6, n. 2, pp. 154-175.
<https://doi.org/10.1080/21670811.2017.1345645>

Bakshy, E.; Messing, S.; Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, v. 348, n. 6239, pp. 1130-1132.
<https://doi.org/10.1126/science.aaa1160>

Benkler, Y.; Faris, R.; Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press. ISBN: 978 0 190923624

Bommasani, R.; Hudson, Drew A.; Adeli, Ehsan; Altman, Russ; Arora, Simran; Von Arx, Sydney; Bernstein, Michael S.; Bohg, Jeannette; Bosselut, Antoine; Brunskill, Emma et al. (2021). On the opportunities and risks of foundation models. Stanford CRFM. *arXiv*.

<https://doi.org/10.48550/arXiv.2108.07258>

Castells, M. (2009). *Communication power*. Oxford University Press. ISBN: 978 0 199567041

Chesney, R.; Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, v. 107, n. 6, pp. 1753-1819.

<https://doi.org/10.2139/ssrn.3213954>

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, v. 43, n. 4, pp. 51-58.

<https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>

European Commission (2018). *A multi-dimensional approach to disinformation*. Publications Office of the European Union.

<https://doi.org/10.2759/739290>

European Union (2024). *Artificial intelligence act*. Official Journal of the European Union.

<https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. Longman. ISBN: 978 0 582219847

Kahneman, D. (2012). *Pensar rápido, pensar despacio*. Debate. ISBN: 978 84 99921705

Kaplan, J.; McCandlish, Sam; Henighan, Tom; Brown, Tom B.; Chess, Benjamin; Child, Rewon; Gray, Scott; Radford, Alec; Wu, Jeffrey; Amodei, Dario (2020). Scaling laws for neural language models. *arXiv*.

<https://doi.org/10.48550/arXiv.2001.08361>

Lazer, D. M. J.; Baum, M. A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; Schudson, M.; Sloman, S. A.; Sunstein, C. R.; Thorson, E. A.; Watts, D. J.; Zittrain, J. L. (2018). The science of fake news. *Science*, v. 359, n. 6380, pp. 1094-1096.

<https://doi.org/10.1126/science.aao2998>

Lewandowsky, S.; Ecker, U. K. H.; Cook, J. (2017). Beyond misinformation: Understanding and coping with the post-truth era. *Journal of Applied Research in Memory and Cognition*, v. 6, n. 4, pp. 353-369.

<https://doi.org/10.1016/j.jarmac.2017.07.008>

McCombs, M. E.; Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, v. 36, n. 2, pp. 176-187.
<https://doi.org/10.1086/267990>

Mirsky, Y.; Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys*, v. 54, n. 1.
<https://doi.org/10.1145/3425780>

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, v. 2, n. 2, pp. 175-220.
<https://doi.org/10.1037/1089-2680.2.2.17>

OECD (2024). *Facts not fakes: Tackling disinformation, strengthening information integrity*. OECD Publishing.
<https://doi.org/10.1787/d909ff7a-en>

Pennycook, G.; Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, v. 188, pp. 39-50.
<https://doi.org/10.1016/j.cognition.2018.06.011>

Sunstein, C. R. (2018). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
<https://doi.org/10.2307/j.ctv8xnhtd>

Tandoc, E. C.; Lim, Z. W.; Ling, R. (2018). Defining “fake news”. *Digital Journalism*, v. 6, n. 2, pp. 137-153.
<https://doi.org/10.1080/21670811.2017.1360143>

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, v. 4, n. 1, pp. 25-29.
<https://doi.org/10.1037/h0071663>

Tversky, A.; Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, v. 185, n. 4157, pp. 1124-1131.
<https://doi.org/10.1126/science.185.4157.1124>

UNESCO (2023). *Guidance for generative AI in education and research*. UNESCO.
<https://doi.org/10.54675/EWZM9535>

Van Dijck, J.; Poell, T.; De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
<https://doi.org/10.1093/oso/9780190889760.001.0001>

Van Dijk, T. A. (1998). *Ideology: A multidisciplinary approach*. Sage Publications. ISBN: 978 0 761956542

Van Dijk, T. A. (2008). *Discourse and power*. Palgrave Macmillan. ISBN: 978 1 4 03996254

Vosoughi, S.; Roy, D.; Aral, S. (2018). The spread of true and false news online. *Science*, v. 359, n. 6380, pp. 1146-1151.
<https://doi.org/10.1126/science.aap9559>

Wardle, C.; Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe. ISBN 978-9287183842.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, v. 9, n. 11, pp. 40-53.
<https://doi.org/10.22215/timreview/1282>

World Economic Forum (2024). *Global risks report 2024*.
<https://www.weforum.org/publications/global-risks-report-2024>

World Health Organization (2020). *Managing the COVID-19 Infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation*.
<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>

Zannettou, S.; Sirivianos, M.; Blackburn, J.; Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality*, v. 11, n. 3.
<https://doi.org/10.1145/3309699>

Zhou, X.; Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, v. 53, n. 5.
<https://doi.org/10.1145/3395046>