

## Confiabilidad en los estudios sobre *fake news*: datasets y métricas<sup>1</sup>

Fake news-*ei buruzko azterlanen fidagarritasuna: dataset-ak eta metrikak*

Reliability in fake news studies: datasets and metrics

Vianny Geraldine Castellanos-Trujillo, Tatiana Hidalgo Marí\*,  
 Patricia Palomares-Sánchez, Raúl Rodríguez-Ferrándiz<sup>2</sup>

Universidad de Alicante

**RESUMEN:** Gran cantidad de estudios empíricos sobre *fake news* emplean como *datasets* repertorios de noticias falsas y de noticias auténticas, al objeto de compararlos. Dichos estudios apelan a menudo a la confiabilidad como factor para caracterizar y evaluar el mensaje, el medio, la fuente y la percepción de la audiencia. A partir de los 50 artículos de metodología empírica más citados en WoS y Scopus (2017–2022) sobre *fake news* y confiabilidad proponemos una metainvestigación donde analizamos los criterios, índices, métricas y bases de datos que declaran los investigadores para justificar su selección tanto de noticias auténticas como falsas.

**PALABRAS CLAVE:** confiabilidad; credibilidad, *fake news*; métricas; *fact-checking*.

**ABSTRACT:** Many empirical studies on disinformation use *fake news datasets* and real news datasets to compare them. Such studies often appeal to reliability (or confidence, trustworthiness, credibility, accuracy) as a factor to characterize and evaluate the message, medium, source and perception of the audience. We selected the 50 most cited articles in WoS and Scopus (2017–2022) on *fake news* and reliability and investigated where the researchers extracted their samples. That is, we propose a meta-research aimed at evaluating how reliable such samples are. We extract and analyze the criteria, indices, metrics, resources and databases that researchers declare to justify their selection of both authentic and false news.

**KEYWORDS:** Reliability, confidence; credibility, *fake news*; metrics; *fact-checking*.

<sup>1</sup> Financiación: Esta investigación ha sido financiada por MCINN/AEI/ 10.13039/501100011033 y la Unión Europea NextGenerationEU/PR TR a través de los proyectos «TRIVIAL: Recursos tecnológicos para el análisis viral inteligente a través de Procesamiento del Lenguaje Natural» (PID2021-122263OB-C22) y «SocialTrust: Análisis de la confiabilidad en medios digitales» (PDC2022-133146-C22).

<sup>2</sup> A los autores ya indicados se suma la participación de Cande Sánchez-Olmos.

\* **Correspondencia a / Corresponding author:** Tatiana Hidalgo Marí. Universidad de Alicante. Ctra. Sant Vicent del Raspeig, s/n. Sant Vicent del Raspeig (Alicante) – [tatiana.hidalgo@ua.es](mailto:tatiana.hidalgo@ua.es) – <https://orcid.org/0000-0003-4599-5876>

**Cómo citar / How to cite:** Castellanos-Trujillo, Vianny Geraldine; Hidalgo Marí, Tatiana; Palomares-Sánchez, Patricia; Rodríguez-Ferrándiz, Raúl (2024). «Confiabilidad en los estudios sobre *fake news*: datasets y métricas», *Zer*, 29(56), 87-109. (<https://doi.org/10.1387/zer.26185>).

Recibido: 31 marzo, 2024; aceptado: 15 mayo, 2024.

ISSN 1137-1102 - eISSN 1989-631X / © UPV/EHU Press



Esta obra está bajo una Licencia  
 Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional

## Introducción

El presente trabajo tiene como objetivo evaluar si el concepto de confiabilidad en la información es una variable en las investigaciones que han trabajado sobre *fake news* y desinformación, y cómo se ha abordado metodológicamente en la investigación empírica. En otras palabras, nos proponemos hacer una metainvestigación para conocer hasta qué punto los textos académicos de alto impacto sobre desinformación que mencionan la confiabilidad (en alguna de las formas en que puede expresarse en inglés: *reliability*, *confidence*, *trustworthiness*, *credibility*, *accuracy*) emplean criterios vinculados a dicha variable para seleccionar sus *datasets*.

Buena parte de la investigación empírica sobre *fake news* se centra en el análisis de la falsedad, al objeto de detectarla (Shu *et al.*, 2017; Viviani & Pasi, 2021; Zhou & Zafarani, 2021) y de ahí prevenir su publicación y viralización (Del Vicario *et al.*, 2017), contrarrestar sus efectos (Roozenbeek & van der Linden, 2019; Maertens *et al.*, 2020) o evaluar la efectividad del *fact-checking* y las alertas (Pennycook *et al.*, 2020; Clayton *et al.*, 2020; Guess, Nyhan & Reifler, 2018; Barrera *et al.*, 2018; Nyhan, 2020).

Otra línea muy consistente de investigaciones analiza qué capacidad real tiene la audiencia para detectar las noticias falsas (Casero Ripollés, 2020), en ocasiones en comparación con su capacidad percibida (Lyons *et al.*, 2021), incluso en colectivos especialmente formados (Palau-Sampio *et al.*, 2022), así como qué factores influyen en esa capacidad: habilidades cognitivas, disposición actitudinal, sesgo político de la noticia, del medio y del informante, tasa de viralidad de la noticia en redes (*endorsement-based heuristics*), reputación del medio en concreto, de la plataforma y de quien comparte la noticia, etc. (Pennycook *et al.*, 2020, 2021; Sterrett *et al.*, 2019; Lee, Gil de Zúñiga, & Munger, 2023; Luo, Hancock & Markowitz, 2022; Shin, Wang & Lu, 2022).

También se ha investigado qué tasa de compartibilidad tienen las noticias falsas con respecto a las verdaderas (Vosoughi, Roy & Aral, 2018) y qué razones mueven a compartir unas y otras (Duffy, Tandoc, Ling, 2019; Osmundsen *et al.*, 2021), así como qué grado de conciencia de la falsedad hay en quien comparte noticias falsas: ignorancia de la falsedad o inatención, partidismo que elude la reflexión sobre la falsedad, disrupción, etc. (Pennycook *et al.*, 2021; Osmundsen *et al.*, 2021; Petersen, Osmundsen, Arceneaux, 2022).

Todas estas investigaciones empíricas implican procesar grandes conjuntos de datos de noticias consideradas falsas y compararlos con otros conjuntos de noticias auténticas. Dado que los investigadores que abordan empíricamente el fenómeno de la desinformación se proveen de dichas bases de datos, sería deseable que no solo de-

claren de dónde extraen unas y otras, sino que recurran a repertorios de unas y otras suficientemente acreditados, que hayan sido validados por instituciones que miden la calidad de la información y/o sometidos a métricas, y que resulten por tanto confiables, tanto en un caso como en otro (es decir, *fact-checkers* acreditados en la detección y desmentido de *fake news* y *legacy media* acreditados en la publicación de informaciones contrastadas).

Partimos de la premisa de que la metodología empleada para recopilar y analizar la información debe ser transparente y replicable, lo que requiere un enfoque meticuloso en la selección de muestras, la metodología de análisis, y la verificación de las fuentes de datos. En este contexto, la relevancia de basarse en muestras empíricas radica en su capacidad para proporcionar evidencia objetiva de los fenómenos estudiados y garantizar que los hallazgos y conclusiones derivados de la investigación puedan ser considerados fidedignos y representativos. El rigor de la investigación se fundamenta en la capacidad de obtener resultados consistentes a lo largo del tiempo y en diferentes estudios que abordan el mismo tema, lo cual es esencial para construir un cuerpo de conocimiento coherente y robusto, en nuestro caso sobre la desinformación y las *fake news*. Esto también implica la transparencia en la metodología de investigación y en los criterios utilizados para definir y clasificar la información como falsa o verdadera, así como la justificación de la selección de las fuentes de datos primarios.

La confiabilidad de las muestras utilizadas en la investigación sobre desinformación y *fake news* es fundamental para garantizar la validez y la aplicabilidad de los resultados, dado que las conclusiones extraídas de estas investigaciones impactan significativamente en las estrategias posteriores para combatir la desinformación. Esto implica no solo la precisión de los datos recopilados, sino también la consideración de sesgos potenciales en la recopilación y presentación de datos. En definitiva, esta precisión metodológica es indispensable no solo para avanzar en la comprensión académica del impacto de las *fake news*, sino también para informar eficazmente las políticas públicas y las estrategias de mitigación (*fact-checking*, Clayton *et al.*, 2020; alfabetización mediática, Sádaba y Salaverría, 2023; inoculación, Roozenbeek & Van der Sanden, 2019) capaces de enfrentar los retos que plantea la desinformación en la era digital y diseñadas para contrarrestar su propagación y efectos nocivos en la sociedad.

## 1. Confiabilidad y credibilidad como factores de la calidad informativa

Confianza y credibilidad son dos términos interrelacionados, ambos de carácter multidimensional, y aplicados a la información se emplean a menudo como sinónimos (Serrano Puche, 2017). Aquí usaremos *confiabilidad* como etiqueta que pretende

en cierto modo reunirlos a los dos y que traduce un conjunto de términos en inglés que presentan algunos matices diferenciales, pero que básicamente se refieren a una única cualidad informativa: *reliability*, *confidence*, *trustworthiness*, *credibility*, *accuracy*.

La confiabilidad tanto de los medios informativos como de los weblogs y las publicaciones en redes sociales tiene que ver con la información de calidad, y es susceptible de ser evaluada (Roberts, 2010). Un medio confiable y creíble es aquel que cumple criterios periodísticos como la reputación, la transparencia y la información ajustada a los hechos (McQuail, 1992). Medir la credibilidad y hacer juicios razonados sobre ella recae en terceras partes: instituciones, observatorios, *fact-checkers* o la propia audiencia (Truong, Allen & Menczer, 2024).

La confianza se ha considerado un componente importante de la credibilidad (Hovland y Weiss, 1951), pero también hay abordajes según los cuales la credibilidad sería un «subfenómeno de la confianza» de carácter más objetivo (Serrano Puche, 2017). Definida como el grado en que un individuo percibe un estímulo como creíble, la credibilidad se ha explicado como una función de tres dimensiones: mensaje, medio y fuente (Metzger *et al.*, 2003). El mensaje se refiere al contenido de un proceso de comunicación, el medio se refiere a la plataforma a través de la cual se envió, y la fuente se refiere a la fuente real o percibida del contenido. Aunque la investigación sobre la credibilidad se centró inicialmente en la credibilidad de la fuente en la comunicación interpersonal, se amplió a medida que los investigadores de la comunicación de masas también examinaron la credibilidad de los medios (Metzger *et al.*, 2003). Sin embargo, los mensajes, tanto en el contexto de la comunicación interpersonal como en el de la comunicación de masas, también tienen características inherentes que los hacen más o menos creíbles (Metzger *et al.*, 2003). Por ello, los estudios también se han centrado en examinar «el grado en que una audiencia cree en un mensaje» (Roberts, 2010, p. 45), como estudio independiente del medio y de la fuente. También se ha explicado para referirse al «juicio de un individuo sobre la veracidad del contenido de la comunicación» (Appelman y Sundar, 2015, p. 63).

La aplicación de criterios de confiabilidad a la información que circula por Internet (Metzger *et al.*, 2003), y en particular a través de redes sociales (Metzger & Flanagin, 2015; Metzger, Flanagin & Medders, 2010; Flanagin, Winter & Metzger, 2020) se ha ido volviendo más y más urgente y también compleja, conforme una parte creciente de la audiencia declaraba informarse de manera prioritaria o exclusiva a través de estas plataformas, y al tiempo reconocía —paradójicamente— desconfiar de esas fuentes más que de los medios tradicionales (Mitchell *et al.*, 2020, Shearer, & Mitchell, 2021).

La confiabilidad, tal y como la entendemos aquí (*accuracy*, *trustworthiness*, *credibility*, *reliability*, *confidence*), forma parte de la dimensión más *intrínseca* de la Information

Quality (IQ). Pero dado que la IQ es una categoría multidimensional, esa dimensión se añade (pero también se interrelaciona) con otras tres: la dimensión *contextual* (relevancia, completez, actualidad, solvencia de la fuente, utilidad), la dimensión *representacional* (claridad, concisión, no ambigüedad, legibilidad, comprensibilidad) y la dimensión de la *accesibilidad* (disponibilidad, usabilidad, seguridad) (Lee *et al.*, 2002, p. 134; Floridi & Illari, 2014).

Además, cuando hablamos de medir la confiabilidad podemos referirnos tanto a un abordaje *ontológico*, es decir, las noticias como mensaje, como contenido, y ello a través de parámetros que exigirán a su vez aproximaciones más cualitativas o datos cuantificables, o bien la confiabilidad *percibida* por la audiencia, en general a partir de encuestas, así como por colectivos cualificados (profesionales del periodismo, Lee *et al.*, 2002, 136). Un tercer ángulo de incidencia tendría que ver con la confiabilidad medida a partir del análisis de las instituciones e infraestructuras de la información (sistema de medios, gobernanza de medios), que determinan el marco empresarial y regulatorio (Lawrence & Napoli, 2023, p. 8).

Un problema difícil resolver sucede cuando esas tres formas de abordar el concepto de la calidad informativa, y por tanto de la confiabilidad de la información, son discrepantes. En particular cuando la percepción de la calidad informativa por las audiencias no coincide con la definición teórica y los parámetros que han determinado los académicos para conceptualizarla y medirla, muy especialmente cuando hablamos de noticias en medios digitales<sup>3</sup>.

Un asunto vinculado es precisamente qué contenidos online son considerados noticias, es decir, cuál es no tanto su *newsworthiness* (qué acontecimientos son más propensos a convertirse en noticias, por el tema, el contexto, la agenda, Boukes, Jo-

---

<sup>3</sup> Uno de los textos más influyentes sobre la propagación de las noticias falsas en Twitter (Vosoughi, Roy & Aral, 2018) plantea precisamente esta cuestión, al reflexionar sobre el hecho de que la polarización en la sociedad norteamericana es tan grande que los medios que aparecen mejor calificados en las métricas e índices al uso confeccionados por los expertos no alcanzan en ningún caso una confiabilidad generalizada entre la ciudadanía. Hasta el punto de que, según un estudio del Pew Research en 2014, las fuentes que los estadounidenses «consistentemente conservadores» y «mayormente conservadores» encuentran confiables (Fox News, por ejemplo) son exactamente las que los estadounidenses «consistentemente liberales» y «mayormente liberales» encuentran poco confiables, y al revés (CNN o *The Washington Post*, en este caso). De esta manera, los autores afirman: «Given this evidence, we do not see how a scientific study could remain objective and take a position on which sources are «reliable» and which are not. Instead, to get at the difference between true and false news, we feel it is imperative to focus on which stories (from any source) have been verified as true or false by multiple independent fact checking organizations». Es decir, plantean recurrir a la confiabilidad de cada noticia en sí, verificada unánimemente por un conjunto de organizaciones independientes, más que la del medio (cfr. Supplementary Materials).

nes & Vliegenhardt, 2022) sino lo que se ha llamado su *news-ness*, definida como «the extent to which audiences characterize specific media content as news» (Edgerly & Vraga, 2020, p. 808).

Reflexionar sobre la calidad de la información a partir de la confiabilidad (y de otros parámetros asociados) supone en cierta medida abordar el problema de la desinformación desde el polo opuesto al habitual, pero complementario (Lawrence, Napoli, 2023, p. 5): no tanto detectar, prevenir o bloquear los efectos de la baja calidad informativa, sino promover en positivo la reclamación de alta calidad informativa por parte de instituciones, responsables de medios y usuarios, así como estimular en las plataformas y en los medios un diseño y unos protocolos que exijan el cumplimiento de estándares de calidad elevados.

## 2. Material y metodología

El presente trabajo es una metainvestigación que explora la producción científica en un ámbito bastante concreto (*fake news* y confiabilidad), y que recopila contribuciones indexadas en dos reputadas bases de datos que evalúan la calidad editorial (Web of Science y Scopus). La finalidad del trabajo es seleccionar una muestra a partir de las contribuciones más citadas y diseccionar la metodología empleada en los trabajos de corte empírico.

Para ello se ha diseñado una metodología de búsqueda bibliométrica específica, creada *ad hoc*, utilizando los términos de búsqueda «fake news», como elemento principal y por otro uno de los siguientes términos: «Accuracy», «Trustworthiness», «Confidence», Reliability) y «Credibility».

La búsqueda se realizó el 27 de marzo de 2023 en las bases de datos WoS (todas las bases de datos incluidas, no solo en el WoS Core Content) y Scopus. Los términos de búsqueda en ambas bases fueron «fake news» por un lado y por otro uno de los siguientes términos: «Accuracy», «Trustworthiness», «Confidence», «Reliability» y «Credibility». Se buscaron dichos términos en TÍTULO, KEYWORDS y ABSTRACT (la búsqueda en esos tres lugares aparece agrupada como «Topics» en WoS), sin acotar temporalmente la búsqueda.

El resultado inicial fueron 1.756 referencias encontradas en Scopus y 1.738 en WoS, lo cual habla de una coincidencia elevada, desde un punto de vista puramente cuantitativo, en el volumen de referencias rastreadas. Procedimos a limitar la búsqueda a Artículos, Libros y Capítulos de Libros, sin acotarla temporalmente. Los resultados arrojaron una muestra más reducida, de 883 referencias en Scopus y de 977 en WoS. Ordenamos los resultados de ambas por el total de citas recibidas. Seleccionamos las 50 más citadas de ambos listados, obteniendo un total de 100 textos.

Ordenamos de mayor a menor esos textos por citas recibidas, eliminando los duplicados. Para ello nos quedamos con el registro en la base que tiene una cifra más elevada de citas recibidas. Dado que no había una coincidencia exacta entre los cincuenta más citados de una base y de la otra, el resultado de cruzar ambas y eliminar los duplicados arrojaba 60 textos resultantes. De ellos, procedimos de nuevo a seleccionar los 50 textos con más citas.

Del resultado de esta operación, resultó que había 29 referencias en las dos bases (58%), mientras las otras 21 procedían de una o de la otra. Y de ellas, 16 en exclusiva de WoS (32%) y 5 en exclusiva de Scopus (10%). Salvo en un caso, los motores de búsqueda devolvieron más resultados de citas en Scopus (29 artículos presentes en ambas más citados en Scopus que en WoS) que en WoS (1), aunque el texto más citado en términos absolutos estaba solo en WoS (Allcott & Gentzkow, 2017).

Ante esta discrepancia entre las bases de datos<sup>4</sup>, tomamos los textos que solo aparecían en una de ellas y lo buscamos de forma manual en la otra. De esa búsqueda manual resultó que, de los 21 textos solo presentes en una de ellas 18 estaban también en la otra. Solo en 3 casos la referencia estaba en Scopus pero no en WoS. Así pues, obtuvimos 47 registros coincidentes (29 en la primera búsqueda, y 18 en la segunda) de los 50 totales, es decir, un 94% de coincidencia. Se confirmó, por tanto, la pertinencia de usar dos bases de datos y cruzar los resultados de ambas para enriquecer y fortalecer la confiabilidad de nuestra propia muestra.

En vistas al análisis cuantitativo, creamos una tabla de Excel *ad hoc* y sometimos a análisis cada uno de los elementos muestrales. Entre las distintas variables, una de las columnas preliminares consistía en determinar si el artículo era teórico o empírico, y si era de estos últimos establecer, en su caso, cuántos estudios empíricos distintos contenía. De ese análisis resultó que 38 de los 50 artículos eran empíricos, y en total contenían 59 estudios distintos, toda vez que ocho de los artículos contenían más de un estudio empírico.

En la siguiente tabla resumimos la metodología de la selección de la muestra llevada a cabo.

---

<sup>4</sup> En cuanto a las razones de la indexación discrepante, podemos aventurar que quizá se deba a que WoS-Clarivate y Scopus emplean para indexar los textos no solo las *keywords* que proporcionan los autores, sino también unas *keywords* determinadas algorítmicamente por la propia base de datos (Keywords Plus, Scopus indexed keywords), lo cual hace que las búsquedas automáticas puedan divergir.

TABLA 1  
Metodología

	Bases de datos académicas	
	Scopus	WoS (todas las bases)
Búsqueda inicial	n = 1756	n = 1738
Acotada a artículos, libros y capítulos de libros	n = 883	n = 977
Selección 50 más citados	50	50
	100	
Eliminación duplicados	60	
Presencia en ambas bases	33 (55,0%)	
Presencia en una u otra	10 (16,6%)	17 (28,3%)
Nueva selección top 50	50	
Presencia en ambas bases	29 (58%)	
Presencia una u otra	5 (10%)	16 (32%)
Búsqueda manual	50	
Presencia en ambas bases	47 (94%)	
Presencia en una u otra	3 (6%)	0 (0%)
Muestra definitiva	50	
Textos teóricos	11 (22%)	
Textos empíricos	39 (78%)	
Total estudios empíricos	59	

Fuente: elaboración propia.

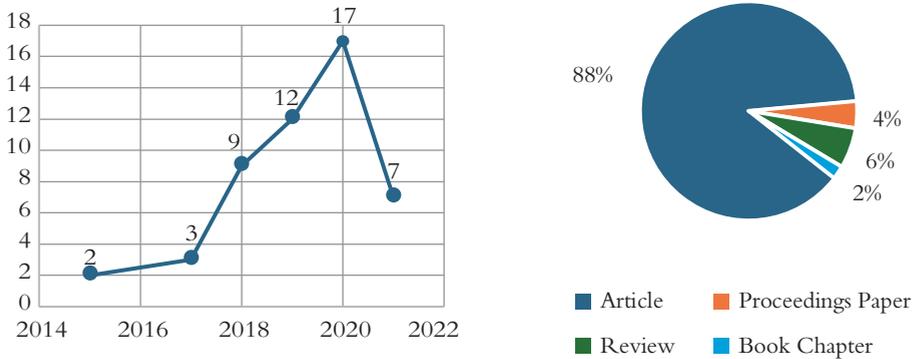
### 3. Resultados

#### 3.1. RADIOGRAFÍA DE LOS ARTÍCULOS DE LA MUESTRA

La búsqueda bibliográfica se basa en los 50 trabajos académicos indexados en Web of Science y Scopus con mayor citación en base al algoritmo: *TITLE-ABS-*

KEY («Fake News») AND TITLE-ABS-KEY («Accuracy» AND «Trustworthiness» AND «Confidence» AND «Reliability»AND «Credibility»). Esta búsqueda, justificada en la metodología, está compuesta, en su mayoría por artículos científicos (88%), seguida muy de lejos de un 6% de reviews y un único capítulo de libro.

En lo que a la distribución temporal se refiere, destaca que, a pesar de que no existe filtro temporal en nuestra búsqueda, los primeros casos aparecen en el año 2015. En concreto, se trata de un artículo científico y una *review* publicados por el equipo investigador formado por Chen, Conroy y Rubin, en los que se analiza, bajo la etiqueta de credibilidad, la proliferación de las noticias falsas desde una aproximación teórica. La franja con mayor volumen de trabajos indexados se localiza en los años 2019-2020, donde se enmarcan el 58% de los trabajos (29 casos).



Fuente: elaboración propia.

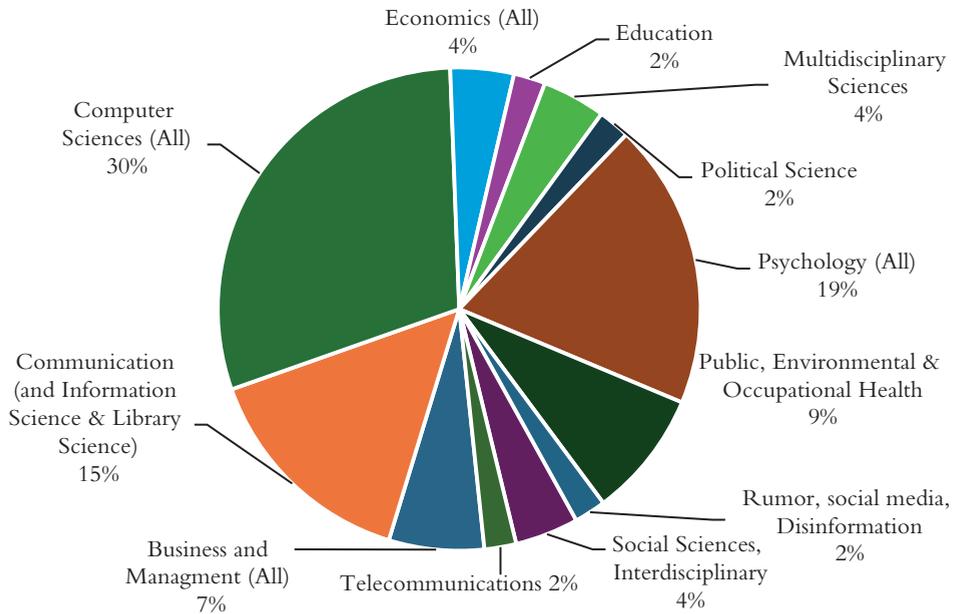
FIGURA 1  
Radiografía de los trabajos

Al analizar las 325 palabras clave indexadas por los autores, calculamos un promedio de 6,5 palabras por artículo analizado. Este promedio se adecua al estándar de exigencia habitual de las revistas científicas, lo que indica una coherencia metodológica en la selección de palabras clave que facilita la indexación y la búsqueda de investigaciones relevantes en este campo. Dentro de este promedio, la palabra más utilizada es el compuesto social-media, que aparece en el 34% de los artículos, lo que subraya la importancia de las plataformas de redes sociales en la difusión de la desinformación, actuando como canales significativos para la propagación de *fake news*. A continuación, le siguen *fake news* (28%), *news* (27%), *misinformation* (24%), *news* y *learning* (14% cada una), *credibility* (10%) o *theory*, *machine* e *information* (8% cada una). La presencia significativa de estos términos en la indexación subraya el enfoque de las investigaciones y el interés emergente por perspectivas teóricas en el estudio de la desinformación.



flejando la necesidad de abordar la desinformación no solo desde un punto de vista tecnológico, sino también considerando sus impactos psicológicos y socioculturales.

Resulta curioso cómo estos trabajos, enfocados en el análisis de la confiabilidad, se alejan de disciplinas como las Ciencias políticas o la Educación. Merece la pena destacar la existencia de una categoría de clasificación temática denominada «rumor, social media and desinformation» que, a pesar de que solo representa un 2% del total de la muestra, apunta a una nueva tendencia temática que ha sido incluida por WoS. Esta tendencia sugiere un esfuerzo por parte de la comunidad académica y de los índices de referencia como WoS por adaptarse a las dinámicas cambiantes del campo de estudio, reconociendo la necesidad de categorías más específicas que reflejen las particularidades de la desinformación en la era digital.



Fuente: elaboración propia.

FIGURA 3  
Ámbitos temáticos de WoS

Si comentamos las palabras clave utilizadas para la búsqueda, detectamos que el término que más artículos ha arrojado al estudio ha sido *accuracy*, presente en el 42% de los *papers* (21 casos), seguido del término *credibility*, que aparece en el 38% de los artículos (17 casos). Los términos *reliability* (18%, 9 casos), *trustworthiness* (4%, 2 casos) y *confidence* (2%, 1 caso) no adquieren demasiada importancia en lo que a

los términos de búsqueda se refiere, a pesar de ser los conceptos clave de la búsqueda inicial.

La distribución temática de los estudios seleccionados sobre desinformación revela un panorama diverso y complejo, donde la agenda política se posiciona como el tema predominante, estando presente en el 46% de los casos analizados. Este enfoque en la agenda política subraya la urgencia de comprender cómo la desinformación puede influir en las actitudes políticas y en el comportamiento electoral, así como en la confianza en las instituciones públicas. El tema de la salud ocupa el segundo lugar en términos de frecuencia, estando presente en el 23% de los casos. La irrupción de la pandemia por COVID-19 ha sido un catalizador para el estudio de la desinformación en este ámbito, representando el 15% del total de casos. En este sentido, la crisis sanitaria global ha puesto de manifiesto la velocidad y el peligro de la difusión de información falsa o engañosa relacionada con la enfermedad, los tratamientos, las vacunas y las medidas de prevención. Además, se detectan otras temáticas relacionadas con la agenda de la salud, que no se relacionan directamente con la crisis del coronavirus (8%). Los temas sociales apenas aparecen en un 10% de los trabajos analizados y, de nuevo, nos encontramos con una gran dispersión en la clasificación temática recogida mediante la etiqueta «otros casos», que alberga el 25% del total. La diversidad de temas subraya la omnipresencia de la desinformación y la necesidad de enfoques multidisciplinarios para su estudio y mitigación.

Centrándonos en los materiales y métodos empleados por los investigadores podemos decir que la metodología más repetida es la empírica, presente en el 78% de los textos analizados, frente al abordaje teórico o especulativo a partir de fuentes bibliográficas y documentales, y no datos obtenidos de primera mano, que se da en el 22% de los casos (ver punto 4.3, donde se desarrolla la metodología empírica). La técnica de investigación más recurrente en los artículos analizados es la aplicación de encuestas y cuestionarios, que puede apreciarse en el 74% de los textos seleccionados, si bien es cierto que es común que se combinen distintas metodologías.

### 3.2. DATASETS EMPLEADOS EN LOS ESTUDIOS EMPÍRICOS

La práctica de utilizar *fact-checkers* para identificar *fake news* (FN) y recurrir a medios de comunicación tradicionales (*legacy media*) para seleccionar *real news* (RN) en estudios de metodología empírica es una tendencia observada en la mayoría de los artículos empíricos sobre desinformación contemplados. Este enfoque implica una delegación implícita de la tarea de determinar la veracidad de las noticias a estas entidades externas, asumiendo que los medios tradicionales son inherentemente confiables para las RN y que los *fact-checkers* proporcionan una fuente autorizada para identificar las FN. Sin embargo, esta práctica también destaca una laguna metodológica: la ausencia de justificación basada en datos sobre la confiabilidad tanto de me-

dios como de *fact-checkers*, en virtud de índices, métricas, sellos de calidad, etc. otorgados por terceras partes independientes que evalúen su confiabilidad.

En las nubes de palabras que siguen representamos los medios más empleados por los artículos empíricos de nuestra muestra para recopilar sus RN (figura 4) y las páginas desinformadoras más citadas en dichos artículos, por ser aquellas cuyas noticias más frecuentemente desmienten los *fact-checkers* a los que recurren los investigadores (figura 5), que aparecen a su vez en la Figura 6.



Fuente: elaboración propia.

FIGURA 4  
Fuentes específicas de Real News

Como se puede observar en la nube de palabras (Figura 4), medios de comunicación como NPR (National Public Radio), *The Washington Post*, *The New York Times*, *The Guardian*, *Fox News* o *The Wall Street Journal* destacan prominentemente, dado que son utilizados con mayor frecuencia como referentes de información veraz en los estudios empíricos. Estos medios, caracterizados por su reputación de periodismo de calidad, parecen ser considerados por la comunidad académica como bastiones de noticias confiables. La representación visual también incluye otros nombres significativos como CNN, BBC y *The Washington Times*, que, aunque en menor medida, indican una recurrencia notable en su uso como fuentes de RN. Se detecta la presencia de una amplia gama de medios, desde aquellos con inclinaciones ideológicas diversas hasta los especializados en negocios como Bloomberg y Fortune, lo que refleja un espectro amplio de fuentes consideradas confiables por los investiga-





### 3.3. ANÁLISIS DE LA JUSTIFICACIÓN DE LOS DATASETS UTILIZADOS

Como decíamos arriba, según el enfoque metodológico utilizado por los autores el 78% de los trabajos que conforman la muestra (39 casos) se basan en una metodología empírica, frente a un 22% de casos que se focalizan en las aproximaciones teóricas (revisiones bibliográficas, a conveniencia o sistemáticas, abordajes especulativos, etc.).

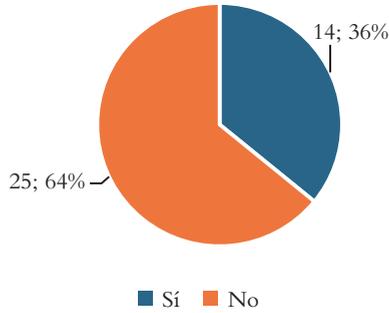
Si nos enfocamos en los estudios empíricos, que son los que más nos interesan, constatamos que de los 39 textos, 32 de ellos no emplean ningún tipo de criterio a la hora de seleccionar sus muestras, sea de RN de FN o ambas (el 82%). Es decir, solo indican de dónde los han extraído (en el caso más repetido de la muestra: Snopes para las FN y *The Washington Post* para las RN), sin apelación a índices o recursos que midan y evalúen la confiabilidad (sea mediante rankings o sellos de calidad). Es cierto que algunos artículos empíricos (p. ej. #28) mencionan en sus estados de la cuestión recursos que evalúan la credibilidad de mensajes en Twitter o de declaraciones políticas, y rastrean o trazan la viralidad de los rumores falsos en la red (CredBank, TwitterTrails, RumorLens, The Emergent), pero no los emplean, como tampoco el resto de artículos.

Solo 7 de los 39 artículos (el 18%: #1, #17, #27, 31#, #36, #37, #43) mencionan y emplean índices, rankings o repertorios que certifican la calidad, y que provienen de instituciones, observatorios o asociaciones (Pew Research Center, Poynter Institute, Reporters without Borders) o aluden a métricas que garantizan la confiabilidad de los *datasets*, u ofrecen índices ponderados del sesgo de las fuentes (MediaBias). Si de los 39 artículos descartamos aquellos que no manejan *datasets* de FN o de RN y FN, comparándolos, o que emplean noticias construidas *ad hoc* para la investigación, la muestra se reduce a 33, pero aun así esos 7 representarían tan solo el 21%.

Además, de esos 7 casos, solo en 3 de ellos las métricas o los índices de confiabilidad se emplean tanto para los repertorios de RN como de FN (al menos una métrica para cada *dataset*, o una común para ambos tipos de repertorios). Son sin embargo 23 los artículos de la muestra de textos empíricos los que, por la naturaleza de su investigación (comparación explícita entre RN y FN, sin fabricación *ad hoc* de noticias para someterlas a los informantes), precisarían recurrir a índices de confiabilidad en ambos repertorios (3 de 23, es decir, el 13%).

En cuanto a los artículos que solo analizan repertorios de FN, son 10 en total (de nuevo descartando aquellos que optan por fabricar *fake news ad hoc* para someterlas a los informantes), pero ninguno de ellos mencionan ni emplean repertorios de medios confiables (en este caso, desconfiables), ni bases de datos específicas donde terceras partes dejan constancia de aquellos *sites* manifiestamente desinformadores,

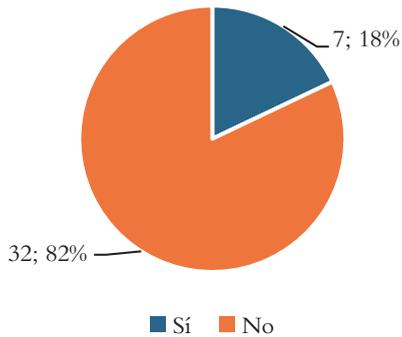
por el volumen de noticias publicadas por ellos que han sido desmentidas o por criterios de relacionados con la calidad periodística, la transparencia, etc.



Fuente: elaboración propia.

FIGURA 7

#### Artículos que emplean métricas de FN y RN para certificar la calidad



Fuente: elaboración propia.

FIGURA 8

#### Artículos que comparan repertorios de FN con repertorios RN para elegir sus muestras

Resulta sintomático que el artículo que alude a más métricas e índices a los que recurrir para verificar la confiabilidad de las fuentes sea precisamente uno teórico, de revisión bibliográfica (Zhou & Zafarani, 2020). Se mencionan, entre otros, *NewsGuard*, índice confeccionado por expertos (analistas, periodistas, editores de varios países) que evalúa en torno a 7.000 webs de noticias de todo el mundo según nueve variables distintas para atribuir credibilidad, *MediaRank*, ránking de medios confiables evaluados por sus pares y *MediaBias/Fact Check*, indicador del sesgo político relativo de cada medio. Dichos recursos apelan a variables que garantizan la

fiabilidad, como la reputación *online* y entre los pares, la medida del sesgo, la declaración de la propiedad, la eventual presión financiera y el conflicto de intereses, la autoría de los textos y de otros contenidos, la transparencia, el recurso a rectificaciones o fe de erratas, la distinción clara entre información y opinión y entre ambas y publicidad, la evitación de titulares engañosos, etc.

Ahora bien, ninguno de esos recursos es mencionado ni menos empleado en ninguno de los 39 textos de metodología empírica.

Son variados y en general altamente fiables tanto estos tres recursos como otros, entre los que podemos mencionar *The Trust Project*, *The Journalism Trust Initiative* (*Reporteros sin fronteras*), *The Global Disinformation Index* (GDI), *The Global Reporting Initiative* (GRI) y *HAZ Foundation's Ranking of Transparency and Governance of Media Companies* (Herrero-Beaumont, 2022), así como *SCImago Media Rankings* (Trillo-Domínguez *et al.*, 2023, 2024), entre otros. Todos ellos emplean métricas o criterios que pueden diferir pero que garantizan altos niveles de calidad en los medios que ocupan las posiciones más altas de la lista, mientras son índices de desconfiabilidad de aquellos otros peor colocados.

#### 4. Discusión y conclusiones

En términos generales podemos decir que los estudios sobre la desinformación, que han crecido significativamente en los últimos años (Alonso García *et al.*, 2020; Park *et al.*, 2020; Righetti, 2021; Tandoc *et al.*, 2018), se caracterizan por abordar temáticas y ámbitos muy dispares, tal y como ha revelado el estudio de palabras clave de los textos de nuestra muestra y hemos resumido sumariamente en la introducción. Además, la diversidad de campos del saber interesados por el fenómeno de la desinformación confirma el impacto en distintas áreas, desde las ciencias de la comunicación y el periodismo a las ciencias de la computación, pasando por la psicología, la sociología, la ciencia política, la bibliometría y la documentación, la filosofía, la economía, etc.

A partir de nuestra muestra de 39 artículos empíricos (que contienen un total de 59 estudios empíricos distintos), que consideramos representativa del conjunto de estudios sobre *fake news* y confiabilidad, podemos extraer algunas conclusiones:

1. Con independencia de si la metodología es el análisis de las noticias en sí mismas (análisis de contenido, del discurso o computacional), o bien las encuestas, *focus groups* o entrevistas en profundidad con informantes acerca de dichas noticias, la gran mayoría de los estudios empíricos recurren a repertorios, bien de noticias falsas, bien de noticias auténticas y de *fake news*, en este caso comparando los resultados para extraer conclusiones.

2. Los objetivos de investigación que se han formulado a partir de esos repertorios de real y *fake news* son muy variados, prácticamente abarcan todos los ámbitos que interesan a priori a una investigación de esta naturaleza. Sin pretender la exhaustividad, encontramos desde estudios que miden el grado de penetración (notoriedad) diferencial en un momento concreto de las RN y las FN sobre un asunto, a otros que observan si las FN se diferencian consistentemente, o no, de las RN en cuanto a temas, formatos, valores, sentimientos, emociones, o bien evalúan hasta qué punto es capaz de detectar las FN la audiencia, qué factores cognitivos o actitudinales influyen en esa capacidad, cómo interviene el titular, el medio, la fuente o las *endorsement cues* (la viralidad en redes cuantificada) en la asignación de confiabilidad a una FN, qué relación hay entre la asignación de confiabilidad a una FN y la decisión de compartirla, qué eficacia cabe otorgar a los *fact-checkers* (*warnings, tags*) a la hora de creer en las FN y decidir compartirlas, etc.
3. La gran mayoría de los estudios toman muestras de RN de medios de referencia, y muestras de FN de *fact-checkers*, pero ni en uno ni en otro caso fundamentan ni justifican la obtención de dichas muestras en métricas o índices de confiabilidad.

Tras explorar la complejidad de la investigación sobre la desinformación y las *fake news*, el presente trabajo destaca importantes desafíos metodológicos y éticos. El más relevante de estos desafíos es la necesidad de justificar la credibilidad de sus muestras de análisis. Ello sugiere que, paradójicamente, los estudios sobre confiabilidad de la información no hacen gala de emplear muestras acreditadamente confiables o, al menos, no justifican suficientemente la fiabilidad de dichas muestras.

El uso de rankings o métricas de medios confiables no es una garantía automática de precisión, pero si las audiencias están crecientemente polarizadas a la hora de asignar credibilidad a las fuentes (ver arriba nota 1), el único acuerdo que parece viable es el de recurrir a datos objetivables y poner en valor a las organizaciones, instituciones, observatorios, etc. que hacen esa labor de monitorizar tanto los medios como las noticias y ofrecer métricas razonadas.

De hecho, podríamos apuntar que la investigación empírica sobre desinformación debería realimentar las propias métricas e índices desde su propia experiencia y resultados. Es decir, no solo recurrir a ellas, identificándolas como herramientas de investigación ajustadas y útiles, sino verificar continuamente su validez, señalar sus posibles deficiencias y por tanto contribuir a refinarlas, labor de extrema importancia en un ecosistema de medios, y de fuentes informales de información, cada vez más volátil y complejo.

## Distribución de trabajo y orden de autoría

Los firmantes declaran que las labores en diseño de la investigación, búsqueda de información, obtención y redacción de introducción, metodología, resultados y conclusiones han sido compartidas sin diferencias reseñables, por lo que se ha establecido como criterio de firma el orden alfabético.

## Referencias bibliográficas

- Allcott, H.; Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-236. <https://doi.org/10.1257/jep.31.2.211>
- Alonso García, S., Gómez García, G., Sanz Prieto, M., Moreno Guerrero, A. J., & Rodríguez Jiménez, C. (2020). The impact of term fake news on the scientific community. Scientific performance and mapping in web of science. *Social Sciences*, 9(5), 73. <https://doi.org/10.3390/socsci9050073>
- Appelman, A., & Sundar, S. S. (2016). Measuring Message Credibility: Construction and Validation of an Exclusive Scale. *Journalism & Mass Communication Quarterly*, 93(1), 59-79. <https://doi.org/10.1177/1077699015606057>
- Barrera, O. *et al.* (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182, 104-123. <https://doi.org/10.1016/j.jpubeco.2019.104123>
- Boukes, M., Jones, N. P., & Vliegenthart, R. (2022). Newsworthiness and story prominence: How the presence of news factors relates to upfront position and length of news stories. *Journalism*, 23(1), 98-116. <https://doi.org/10.1177/1464884919899313>
- Casero-Ripollés, A. (2020). Impact of Covid-19 on the media system. Communicative and democratic consequences of news consumption during the outbreak. *El profesional de la información*, 29(2), e290223. <https://doi.org/10.3145/epi.2020.mar.23>
- Clayton, K. Et. Al. (2020). Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media. *Political Behavior*, 42, 1073-1095. <https://doi.org/10.1007/s11109-019-09533-0>
- Del Vicario, M. *et al.* (2019). Polarization and Fake News: Early Warning of Potential Misinformation Targets. *ACM Transactions on the Web*, 13(2) <https://doi.org/10.1145/3316809>
- Duffy, A., Tandoc, E., & Ling, R. (2020). Too good to be true, too good not to share: the social utility of fake news. *Information, Communication & Society*, 23(13), 1965-1979. <https://doi.org/10.1080/1369118X.2019.1623904>
- Edgerly, S. & Vraga, E. K. (2020). That's Not News: Audience Perceptions of «News-ness» and Why It Matters. *Mass Communication and Society*, 23(5), 730-754. <https://doi.org/10.1080/15205436.2020.1729383>

- Flanagin, A.J., Winter, S. & Metzger, M.J. (2020) Making sense of credibility in complex information environments: the role of message sidedness, information source, and thinking styles in credibility evaluation online, *Information, Communication & Society*, 23(7), 1038-1056. <https://doi.org/10.1080/1369118X.2018.1547411>
- Floridi, L., & Illary, Ph. (2014). *The philosophy of information quality*. Springer.
- Guess, A., Nyhan, B., & Reifler, J. (2018). Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign. European Research Council, 9.
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15, 635-650.
- Herrero-Beaumont, E. (2022). Emerging Transparency Systems for News Governance to Protect Media Independence and Credibility in the Digital Infosphere. *Communication Law and Policy*, 27(3-4), 220-249. <https://doi.org/10.1080/10811680.2022.2154071>
- Lawrence, R. G., Napoli, P. M. (eds.) (2023). *News quality in the digital age*. London: Routledge.
- Lee, S., Gil De Zúñiga, H., & Munger, K. (2023). Antecedents and consequences of fake news exposure: a two-panel study on how news use and different indicators of fake news exposure affect media trust. *Human Communication Research*, hqad019. <https://doi.org/10.1093/hcr/hqad019>
- Lee, Y. W., Strong, D.M., Kahn, B.K., & Wang, R.Y. (2002). AIMQ: a methodology for information quality assessment. *Information & Management*, 40, 133-146. [https://doi.org/10.1016/S0378-7206\(02\)00043-5](https://doi.org/10.1016/S0378-7206(02)00043-5)
- Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. *Communication Research*, 49(2), 171-195. <https://doi.org/10.1177/0093650220921321>
- Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., & Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences*, 118(23), e2019527118. <https://doi.org/10.1073/pnas.2019527118>
- Maertens, R., Roozenbeek, J., Basol, M., & Van Der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1-16. <https://doi.org/10.1037/xap0000315>
- Mcquail, D. (1992). *Media Performance: Mass Communication in the Public Interest*. London: Sage.
- Metzger, M. J. et al. (2003). Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment. *Annals of the International Communication Association*, 27(1), 293-335. <https://doi.org/10.1080/23808985.2003.11679029>
- Metzger, M. J., & Flanagin, A. J. (2015). Psychological approaches to credibility assessment online. In S. S. Sundar (ed.), *The handbook of the psychology of communication technology* (pp. 445-466). West Sussex, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118426456.ch20>

- Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60(3), 413-439. <https://doi.org/10.1111/j.1460-2466.2010.01488.x>
- Mitchell, A., Jurkowitz, M., Oliphant, J. B., & Shearer, E. (2020, July 30). Americans Who Mainly Get Their News on Social Media Are Less Engaged, Less Knowledgeable. Pew Research Center. <https://pewrsr.ch/437CXY1>
- Nyhan, B. *et al.* (2020). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on Factual Beliefs and Candidate Favorability. *Political Behavior* 4, pp. 939-960. <https://doi.org/10.1007/s11109-019-09528-x>
- Osmundsen, M. *et al.* (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review*, 115(3), 999-1015. <https://doi.org/10.1017/S0003055421000290>
- Palau-Sampio, D., Carratalá, A., Tarullo, R., & Crisóstomo, P. (2022). Quality recognition as a prescriber against disinformation. *Comunicar*, 72, 59-70. <https://doi.org/10.3916/C72-2022-05>
- Park, A. *et al.* (2020). Understanding Fake News: A Bibliographic Perspective. *Defence Strategic Communications*, 8. <https://doi.org/10.30966/2018.RIGA.8.4>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592, 590-595. <https://doi.org/10.1038/s41586-021-03344-2>
- Pennycook, G., Mcphetres, J., Zhang, Y., Lu, J. G., Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31(7), 770-780. <https://doi.org/10.1177/0956797620939054>
- Petersen M. B. Osmudsen, M., & Arceneaux, K. (2023). The «Need for Chaos» and Motivations to Share Hostile Political Rumors. *American Political Science Review*, 117(4), 1486-1505. <https://doi.org/10.1017/S0003055422001447>
- Righetti, N. (2021). Four years of fake news: A quantitative analysis of the scientific literature. *First Monday*, 26(6-7). <https://doi.org/10.5210/fm.v26i7.11645>
- Roozenbeek, J. & Van Der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communication*, 5, 65. <https://doi.org/10.1057/s41599-019-0279-9>
- Roberts, C. (2010). Correlations among variables in message and messenger credibility scales. *American Behavioral Scientist*, 54(1), 43-56. <https://doi.org/10.1177/0002764210376310>
- Sádaba, C. & Salaverría, R. (2023). Combatir la desinformación con alfabetización mediática: análisis de las tendencias en la Unión Europea. *Revista Latina de Comunicación Social*, 81, 17-33. <https://www.doi.org/10.4185/RLCS-2023-1552>
- Serrano-Puche, J. (2017). Credibilidad y confianza en los medios de comunicación: un panorama de la situación en España. En González, M. y Valderrama, M. (coords.) (2017), *Discursos comunicativos persuasivos hoy* (pp. 427-438). Madrid: Tecnos.

- Shearer, E. & Mitchell, A. (2021, January 12). News use across social media platforms in 2020. Pew Research Center. <https://t.ly/S6tSd>
- Shin, I., Wang L., & Lu, Y. (2022). Twitter and Endorsed (Fake) News: The Influence of Endorsement by Strong Ties, Celebrities, and a User Majority on Credibility of Fake News During the COVID-19 Pandemic. *International Journal of Communication*, 16, 2573-2595.
- Shu, K. *et al.* (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19, 24-36. <https://doi.org/10.1145/3137597.3137600>
- Sterrett, D. *et al.* (2019). Who Shared It?: Deciding What News to Trust on Social Media. *Digital Journalism*, 7(6), 783-801. <https://doi.org/10.1080/21670811.2019.1623702>
- Tandoc *et al.* (2018). Audiences' acts of authentication in the age of fake news: A conceptual framework. *New Media & Society*, 20(8) 2745-2763.
- Trillo-Domínguez, M., Salaverría, R., Codina, L., & De-Moya-Anegón, F. (2023). SCImago Media Rankings (SMR): situation and evolution of the digital reputation of the media worldwide. *Profesional De La información*, 32(5).
- Trillo-Dominguez *et al.* (2024). Digital reputation indicator: A webometric approach for a global ranking of digital media, *Journalism*. <https://doi.org/10.1177/14648849241237647>
- Truong, B.T., Allen, O.M. & Menczer, F. (2024). Account credibility inference based on news-sharing networks. *EPJ Data Sci.* 13(10). <https://doi.org/10.1140/epjds/s13688-024-00450-9>
- Viviani, M., & Pasi, G. (2017). Credibility in social media: opinions, news, and health information—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7. <https://doi.org/10.1002/widm.1209>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359, 1146-1151. <https://www.science.org/doi/10.1126/science.aap9559>
- Zhou, X. & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, 53(5). <https://doi.org/10.1145/3395046>

