

MARÍA DEL MAR RAMÍREZ-ALVARADO
coordinadora

INTELIGENCIA ARTIFICIAL Y DESINFORMACIÓN

Editorial Universidad de Sevilla

INTELIGENCIA ARTIFICIAL Y DESINFORMACIÓN

MARÍA DEL MAR RAMÍREZ-ALVARADO
COORDINADORA

INTELIGENCIA ARTIFICIAL Y DESINFORMACIÓN

 EDITORIAL
UNIVERSIDAD DE SEVILLA

SEVILLA 2025

Colección Abierta (EUS)

Núm.: 62

Serie Monografías_CAA (Consejo Audiovisual de Andalucía)

Núm.: 1

Comité editorial de
la Editorial Universidad de Sevilla:

Araceli López Serena
(Directora)

Elena Leal Abad
(Subdirectora)

Concepción Barrero Rodríguez
Rafael Fernández Chacón
María Gracia García Martín
María del Pópulo Pablo-Romero Gil-Delgado
Manuel Padilla Cruz
Marta Palenque
María Eugenia Petit-Breuilh Sepúlveda
Marina Ramos Serrano
José-Leonardo Ruiz Sánchez
Antonio Tejedor Cabrera

Reservados todos los derechos. Ni la totalidad ni parte de este libro puede reproducirse o transmitirse por ningún procedimiento electrónico o mecánico, incluyendo fotocopia, grabación magnética o cualquier almacenamiento de información y sistema de recuperación, sin permiso escrito del Consejo Audiovisual de Andalucía y de la Editorial Universidad de Sevilla.

Motivo de cubierta: Karina Torres Ciano. Diseño sobre una imagen generada parcialmente con Gemini (Google DeepMind, 2025)

© Consejo Audiovisual de Andalucía 2025
c/ Pagés del Corro, 90B - 41010 Sevilla
Tlfs.: 955 407 310 | 900 159 159
Correo electrónico: info.caa@juntadeandalucia.es
Web: <https://consejoaudiovisualdeandalucia.es>

© Editorial Universidad de Sevilla 2025
c/ Porvenir, 27 - 41013 Sevilla.
Tlfs.: 954 487 447; 954 487 451
Correo electrónico: info-eus@us.es
Web: <https://editorial.us.es>

© De los textos, los autores 2025
ISBN 978-84-472-2708-2
DOI <https://dx.doi.org/10.12795/9788447227082>

Diseño de cubierta, diseño de interior, maquetación e interactividad:
Referencias Cruzadas. referencias.maquetacion@gmail.com

Índice

Prólogo	
Domi del Postigo	9
1 Inteligencia artificial y desinformación: perspectiva desde los organismos europeos reguladores del audiovisual	
María del Mar Ramírez-Alvarado.....	11
2 Inteligencia artificial contra la desinformación: análisis internacionales sobre plataformas de <i>fact-checking</i> iberoamericanas y europeas	
María Sánchez González y Hada Sánchez Gonzales	33
3 La audiencia juvenil ante la desinformación: percepciones y reacciones en torno a la información política	
Amparo Huertas Bailén y Luiz Peres-Neto	47
4 Inteligencia artificial, infoxicación e identificación de la desinformación por parte de jóvenes andaluces	
Aránzazu Román-San-Miguel y Marta Sánchez-Hunt	65
5 Retos y oportunidades de la inteligencia artificial: uso y percepción entre adolescentes, familias y docentes	
Xavier Bringué Sala y Charo Sádaba Chalezquer.....	77
6 El sexismo digital y la inteligencia artificial, otras formas de desigualdad y exclusión	
Lucía Benítez Eyzaguirre	89

7	■ Ética y transparencia: claves del uso responsable de la inteligencia artificial en Radio y Televisión Española (RTVE) Yasir Barceló, Nereida Cea y Sonia Blanco	103
8	■ <i>Humano ex machina</i> : enfoque y ejemplos para la creación de herramientas de inteligencia artificial contra la desinformación Javier Cantón-Correa	117
9	■ Desinformación en contextos tecnológicos y digitales: impacto, marcos legales y estrategias para contrarrestarla Mónica Niño Romero	135
10	■ Creando una inteligencia artificial capaz de analizar piezas audiovisuales Eduardo Lleida Solano, Enma Calvo Olloqui y Luis Deltell Escolar	153
11	■ <i>Deepfakes</i> y desinformación organizacional: una revisión de la literatura Sónia Gomes Gonçalves y Pedro Álvaro Pereira Correia	171
12	■ Cómo investigar la inteligencia artificial: tres estudios de caso Santiago Tejedor y Stephanie Vick	195
	Autorías	213

Prólogo

Domi del Postigo

Presidente Consejo Audiovisual de Andalucía
josef.dominguez.postigo@juntadeandalucia.es

Siempre en abierto. El Conocimiento, sí, en mayúscula, debe fluir como el agua, sin cerramientos, entre la gente. Y, cuando la gente con sed accede al Conocimiento, ya no es solo gente, es ciudadanía libre e informada. Por eso, el Consejo Audiovisual de Andalucía (CAA) apoya el Conocimiento abierto. En esa línea de actuación, el CAA, a través de la Editorial Universidad de Sevilla, edita este libro que compila el conocimiento expuesto por los investigadores participantes en la primera edición de su Seminario de Investigaciones y Estudios del Audiovisual (SeminCAA).

El Seminario de investigadores no es la única actividad ni este libro la única publicación del CAA al respecto. Y es lógico que sea así, dentro del funcionamiento que debe tener una entidad como esta. El CAA es un órgano extraparlamentario independiente que se ocupa por imperativo legal de regular y vigilar, en defensa de la ciudadanía (en especial de los menores y los sectores más vulnerables de la población), los contenidos audiovisuales emitidos por cualquier medio y tecnología. Y lo hace aplicando su potestad de manera directa o advirtiendo, como intermediario con certificado institucional, donde no alcanza su competencia sancionadora, del riesgo o ilegalidad de algunos de esos contenidos, y trasladando ese informe a las entidades andaluzas, estatales y/o europeas competentes, en red. El CAA es socio de la Plataforma de Reguladores del Sector Audiovisual de Iberoamérica (PRAI) y la Plataforma Europea de Reguladores del Audiovisual (EPRA), entre otras asociaciones

supranacionales que aglutinan autoridades audiovisuales institucionales independientes. En España, el Consejo Audiovisual Andaluz forma parte del Grupo de Autoridades de Supervisión para los Servicios Audiovisuales (GASSCA), presidido por la Comisión Nacional de los Mercados y la Competencia (CNMC), el regulador estatal, conformado en cumplimiento de la disposición adicional segunda de la Ley 13/2022, de 7 de julio, General de Comunicación Audiovisual. Como miembro del GASSCA, el CAA lidera el grupo de trabajo sobre Alfabetización Mediática Informativa.

Desde esa ubicación resumida y siguiendo una de sus líneas de actuación, el Consejo puso en marcha el SeminCAA. En esta primera convocatoria del Seminario, se enfocó la presentación de los estudios realizados, cuya selección se defendió luego por sus autores de manera mayoritariamente presencial en el Salón de Plenos del Consejo, en tres asuntos fundamentales: Inteligencia Artificial, Desinformación y la conexión entre ambos (una relación que supone todo un desafío en la actualidad).

En este libro quedan reflejados esos estudios. Por ello, esta publicación del CAA es un granito de arena más, que pretende sumar transferencia de conocimiento en la adecuada relación entre las competencias y materias de esta institución y la sociedad a la que se debe.

También esta publicación es un agradecimiento en sí misma a los investigadores que nos han aportado ese conocimiento en materia audiovisual (que no está hoy relacionado con los altavoces y las pantallas en nuestro contexto digital), así como a las universidades que les acogen.

Inteligencia artificial y desinformación: perspectiva desde los organismos europeos reguladores del audiovisual

María del Mar Ramírez-Alvarado

Consejera del Consejo Audiovisual de Andalucía
mariam.ramirez.alvarado@juntadeandalucia.es

1 | Consideraciones generales

El vertiginoso avance de la inteligencia artificial (IA) ha supuesto un importante cambio de paradigma en el ámbito de la comunicación. Su irrupción y rápida evolución, junto con la extendida transformación digital, está contribuyendo a la generación de beneficios de distinto orden, proporcionando ventajas competitivas. Pero, al mismo tiempo, son diversos los riesgos asociados, en especial en lo que respecta a la desinformación. La falta de transparencia y ética en la aplicación de la IA es también una preocupación generalizada, así como su vinculación con el auge de los discursos de odio, el contenido perjudicial para menores y el impacto en la difusión de noticias falsas y en los procesos electorales.

Este nuevo paradigma ha puesto sobre la palestra el papel que las autoridades reguladoras en materia del audiovisual pueden ejercer. Esta es, sin duda, una excepcional coyuntura que ha impulsado, tanto en España como a nivel europeo e internacional, la necesidad de contar con un marco regulatorio de la IA que esté ajustado a las sociedades democráticas y al Estado de derecho.

La transformación en esta materia, aunada al creciente uso de los servicios digitales, conlleva nuevos riesgos y desafíos para los destinatarios individuales de los correspondientes servicios, para las empresas e instituciones y, en líneas generales, para la sociedad en su conjunto. En este contexto, los organismos reguladores han puesto en marcha iniciativas tendentes a comprender, valorar y evaluar el uso de la IA en el marco de sus funciones de protección de los derechos fundamentales de la ciudadanía en el ámbito audiovisual (ampliado al digital), de la salvaguarda de los menores y colectivos vulnerables, de la igualdad de género y no discriminación, de la libertad de expresión y del pluralismo informativo, entre otros,

Las aportaciones, posibilidades y avances que la IA ofrece son innegables, pero ciertamente resulta fundamental implementar medidas que garanticen su uso seguro. Diversas investigaciones advierten sobre cómo los algoritmos pueden entrenarse con datos que muestran prejuicios y sesgos. A ello se añaden los problemas vinculados a la vulneración de la privacidad y derecho a la intimidad de las personas, así como al acceso de los menores a contenidos inapropiados. Eso, sumado al uso de la IA para la desinformación y la manipulación.

Estamos, por tanto, ante un horizonte nuevo que es necesario comprender y afrontar a través de medidas eficaces. Los datos del último *Barómetro Audiovisual de Andalucía del CAA* muestran, en cuanto al nivel de conocimiento acerca de la IA, que 77,5% de la muestra encuestada tiene un conocimiento nulo (45,8%) o bajo (31,7%) sobre la IA (CAA, 2025: 215). Este dato es llamativo si se cruza, además, con el que demuestra que, más allá del nivel de estudios alcanzados, la mayoría de la población andaluza dice tener un nulo conocimiento acerca de la IA (CAA, 2025: 216). Y es destacable también que, con independencia de la edad y aludiendo a las desventajas de la IA, son elevados los porcentajes que reflejan la creencia de que el uso de la IA reemplazará a los seres humanos (50,5%) (CAA, 2025: 225).

Este cambio de paradigma vinculado a la irrupción de la IA conecta con otra de las grandes preocupaciones de la ciudadanía y es la que tiene que ver con la desinformación. La Comisión Europea define el término como: «información verificablemente falsa o engañosa que se crea, presenta y divulga con fines lucrativos o para engañar deliberadamente a la población, y que puede causar un perjuicio público». El perjuicio público comprende amenazas contra los procesos democráticos políticos y de elaboración de políticas, así como contra los bienes públicos, como la protección de la salud, el medio ambiente o la seguridad de los ciudadanos de la Unión Europea (Comisión Europea, 2018: 3)¹.

1. Es importante destacar que en esta definición de «desinformación» no incluye la publicidad engañosa, los errores de información, la sátira y la parodia ni las noticias y los comentarios claramente identificados como partidistas.

Este capítulo ofrece un panorama de la actividad en materia de IA y desinformación por parte de los organismos reguladores del audiovisual en Europa, plataformas y redes internacionales que los agrupan. Se presentan las estrategias implementadas en el ámbito de la IA para combatir la desinformación, incluyendo la educación mediática y la regulación de plataformas digitales, entre otras. También se mencionan iniciativas para valorar y frenar el impacto negativo de la IA, tanto en la creación de *deepfakes* como en su detección. Además, se examinan las iniciativas de colaboración entre reguladores y la puesta en marcha de un marco legal actualizado.

2 | Organismos reguladores del audiovisual y plataformas asociativas estudiadas

Las que se detallan a continuación son las autoridades reguladoras y plataformas asociativas cuyas actividades han sido consultadas para este estudio. Los organismos reguladores comparten como característica emplazarse en el ámbito europeo, aunque tienen distintos perfiles y alcances. Las asociaciones poseen un carácter más amplio, pero tienen en común que en ellas se integran los reguladores europeos.

- Comisión Nacional de los Mercados y la Competencia (CNMC), España: es un organismo público con personalidad jurídica propia, independiente del Gobierno y sometido al control parlamentario. Entró en funcionamiento en 2013 unificando una pluralidad de organismos reguladores existentes (uno de ellos el Consejo Estatal de Medios Audiovisuales) con el fin de reforzar su independencia, garantizar la seguridad jurídica y la confianza institucional. El objetivo fue el de aportar una visión integradora en regulación y defensa de la competencia para promover la modernización de la economía en beneficio de los consumidores.
- Consejo Audiovisual de Andalucía (CAA), España: desde 2005 es la autoridad independiente que regula el sector audiovisual de Andalucía. Su cometido es velar porque los medios audiovisuales cumplan los derechos y libertades reconocidos en la Constitución y el Estatuto de Autonomía de Andalucía, así como por el cumplimiento de la normativa vigente en materia audiovisual y de publicidad. La Ley 1/2004 de 17 de diciembre da carta de naturaleza a la creación del CAA.
- Consell de l'Audiovisual de Catalunya (CAC): creado por la Ley 2/2000, de 4 de mayo, es la autoridad reguladora autonómica independiente en el ámbito de la comunicación audiovisual pública y privada, con personalidad jurídica pública propia,

actuando con plena independencia del Gobierno de la Generalitat y de las administraciones públicas en el ejercicio de sus funciones.

- Consell del Audiovisual de la Comunitat Valenciana (CACV), España: con carta de naturaleza a través de la Ley 10/2018, de 18 de mayo, aprobada por la Generalitat Valenciana, es la autoridad independiente que vela en esta comunidad autónoma por el cumplimiento de la normativa vigente en materia audiovisual y de publicidad, actuando en interés público para garantizar una información veraz.
- Autorità per le Garanzie nelle Comunicazioni (AGCOM), Italia: creada mediante la Ley 249 del 31 de julio de 1997, AGCOM es la autoridad nacional independiente italiana de regulación de las industrias de la comunicación. Cumple funciones de regulación y de supervisión de los sectores asociados a las telecomunicaciones.
- ARCOM_Autorité de Régulation de la Communication Audiovisuelle et Numérique (Francia): con este nombre se configura en 2022 la integración del Consejo Superior del Audiovisual y la Alta Autoridad para la Difusión de Obras y la Protección de Derechos en Internet mediante la Ley 2021-1382 del 25 de octubre de 2021. Es la autoridad pública independiente de Francia encargada de garantizar la libertad de comunicación y el cumplimiento de las obligaciones legales y reglamentarias en los sectores audiovisual y también digital.
- Entidade Reguladora para a Comunicação Social (ERC), Portugal: creada por la Ley n.º 53/2005 que extinguió a la anterior Alta Autoridade para a Comunicação Social, es el organismo responsable en Portugal de la regulación y supervisión de los medios de comunicación. Dio inicio a su actividad en 2006.
- Die Medienanstalten (Alemania): en Alemania son catorce las autoridades reguladoras de medios. Cada estado federal (*Land*) dispone de su propia autoridad de medios, promoviendo así la diversidad y el pluralismo en el panorama mediático alemán. Se agrupan en la estructura descentralizada de Die Medienanstalten, lo que garantiza la libertad y evita las concentraciones de poder.
- Conseil Supérieur de l'Audiovisuel (CSA) (Federación Valonia-Bruselas): se crea como un órgano consultivo integrado en la administración de la Comunidad Francesa de Bélgica en 1987 con el objetivo de transferir, desde el gobierno federal belga hacia las comunidades lingüísticas, las competencias en materia audiovisual. Más tarde, el Decreto del 24 de julio de 1997 le conferiría poderes de control y sanción, así como independencia institucional

para gestionar en la comunidad francófona su sector audiovisual adaptándolo a sus especificidades.

- OFCOM (Reino Unido): la Office of Communications Act 2002 sentó las bases para la unificación de cinco organismos que existían hasta entonces: la Broadcasting Standards Commission, la Independent Television Commission, la Office of Telecommunications (OFTEL), la Radio Authority y la Radiocommunications Agency. Un año después es la Communications Act 2003 la que confiere al OFCOM funciones de regulación de los servicios de comunicación autoridad, promoción de la competencia, protección de los usuarios y gestión del espacio radioeléctrico. Con el tiempo ha asumido también la función de regulación de contenidos en línea y la protección de los usuarios en el entorno digital.

En cuanto a las plataformas de organismos reguladores:

- European Regulators Group for Audiovisual Media Services (ERGA): el 10 de febrero de 2025 se constituyó oficialmente el European Board for Media Services (Consejo Europeo de Servicios de Medios de Comunicación) en cumplimiento del Reglamento Europeo sobre la Libertad de los Medios de Comunicación del que se hablará a continuación. Desde entonces, ha asumido las funciones la ERGA como órgano independiente. Se trata de una entidad (ahora *Media Board*) que agrupa a las autoridades reguladoras de nacionales en el ámbito de los servicios audiovisuales de la Unión Europea. ERGA fue creada originalmente en 2014 por la Comisión Europea con las funciones de asesorar en la aplicación de la Directiva de Servicios de Comunicación Audiovisual en todos los Estados miembros y de promover la cooperación y el intercambio de experiencias y buenas prácticas².
- European Platform of Regulatory Authorities (EPRA): fundada en 1995, la EPRA surgió debido a la necesidad de incrementar la colaboración entre las autoridades reguladoras en Europa (actualmente está integrada por de 52 autoridades reguladoras de 46 países miembros). Funciona en gran medida como un foro de discusión y de intercambio de información, así como de conocimiento y asesoría en cuanto a la aplicación en Europa de normativas y legislación concernientes a la regulación de medios de comunicación.
- Mediterranean Network of Regulatory Authorities (MNRA) / Réseau des Instances de Régulation Méditerranéennes (RIRM): se

2. Dado que el cambio es reciente, recogemos su actuación como ERGA en este trabajo.

trata de una organización que agrupa a las autoridades reguladoras del ámbito audiovisual de los países de la cuenca mediterránea y que cuenta actualmente con veintisiete miembros. Sus objetivos fundacionales están en la línea de ofrecer un espacio para debate y la discusión de temas de interés común, propiciando el intercambio de información y el desarrollo de investigaciones sobre el audiovisual.

- Plataforma de Reguladores del Sector Audiovisual de Iberoamérica (PRAI): red internacional que funciona desde 2014 (aunque se empezó a gestar en 2010) y que reúne a organismos independientes encargados de la regulación del sector audiovisual en países iberoamericanos. Su principal objetivo es fomentar el intercambio de experiencias y recursos, propiciando la coordinación en acciones conjuntas y la generación de conocimiento y de información académica que sustenten decisiones regulatorias. En la PRAI están integrados doce organismos reguladores o similares de Argentina, Chile, España, Perú, Portugal, Ecuador, México, Colombia y Brasil) y cuenta con dos miembros observadores (Costa Rica y Marruecos).

3. Inteligencia artificial y desinformación: ejes temáticos centrales

El estudio de las iniciativas puestas en marcha (o que han influido en la puesta en marcha...) a lo largo en los dos últimos años (2023-2025) desde los organismos reguladores mencionados y las plataformas asociativas en materia de IA y desinformación ha permitido definir los siguientes tres ejes temáticos en los cuales se desarrollan actuaciones: IA, legislación y estrategias regulatorias; IA, verificación e identificación de contenidos, y efectos de la IA, alfabetización mediática y otras estrategias de prevención

3.1. Inteligencia artificial, legislación y estrategias regulatorias

Si bien el marco normativo vinculado a la IA se encuentra en constante transformación y sometido a continuos cambios legislativos, los organismos reguladores del audiovisual coinciden en su empeño por estudiarlos, comprenderlos y adaptarse en el marco de sus competencias dependiendo de los países. A efectos de este trabajo, mencionaremos las respuestas legislativas recientes a los desafíos y riesgos de la IA en el período estudiado.

3.1.1. Reglamento de Servicios Digitales: Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo, de 19 de octubre de 2022, relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE

Este Reglamento de Servicios Digitales, conocido como DSA (Digital Services Act) se encuentra vigente para los Estados miembros de la Unión Europea desde el 17 de febrero de 2024 y ofrece un marco unificado aplicable a los servicios de la sociedad de la información para la consecución de un entorno digital seguro. Estos servicios digitales comprenden, fundamentalmente, a los intermediarios y las plataformas en línea: mercados en línea, redes sociales, plataformas de intercambio de contenidos, entre otras. En el campo que nos ocupa, obliga a los Estados miembros a designar, al menos, una autoridad encargada de supervisar su aplicación y ejecución, de acuerdo con el Derecho nacional. También es posible encomendar a más de una autoridad el desempeño de determinadas funciones y competencias de supervisión en sectores concretos como el de las comunicaciones electrónicas, los reguladores de medios de comunicación o las autoridades de protección de los consumidores. Así:

Estas autoridades deben contribuir al logro de objetivos del Reglamento, a saber, el correcto funcionamiento del mercado interior de servicios intermediarios en el que las normas armonizadas para un entorno en línea seguro, predecible y fiable establecidas que facilite la innovación, y en particular, las obligaciones de diligencia debida aplicables a las diferentes categorías de prestadores de servicios intermediarios, se supervisen y apliquen realmente con el fin de garantizar la protección efectiva de los derechos fundamentales amparados por la Carta, incluido el principio de protección de los consumidores (Reglamento UE 2022/2065: 109).

La DSA impulsa la colaboración entre las autoridades nacionales competentes y la Comisión Europea para garantizar una aplicación coherente y efectiva de sus disposiciones. Esto implica que los organismos reguladores del audiovisual deben coordinarse con otras entidades reguladoras para supervisar el cumplimiento de las obligaciones impuestas a los prestadores de servicios digital. Tiene, además, implicaciones para los organismos reguladores del audiovisual en el sentido de que obliga a las plataformas a eliminar contenido ilegal, proteger a menores, ser transparentes sobre sus algoritmos y ofrecer herramientas para denunciar contenidos inapropiados.

Por otra parte, ya que se exige a las plataformas en línea y otros intermediarios digitales medidas de transparencia (incluyendo informes sobre moderación de contenidos y la gestión de riesgos sistémicos), los reguladores del audiovisual tienen la responsabilidad de evaluación de los

mismos. También han de velar por que las plataformas implementen mecanismos efectivos para identificar y retirar dichos contenidos ilícitos y la desinformación atendiendo siempre al derecho a la libertad de expresión.

Como coordinadora de servicios digitales en España, la CNMC supervisa y asegura el cumplimiento de la DSA. Esto incluye la supervisión, investigación y sanción de plataformas digitales, con acceso a datos y algoritmos.

3.1.2. Reglamento de inteligencia artificial: Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial³

El conocido de forma resumida como Reglamento de inteligencia artificial entró en vigor en la UE el 1 de agosto de 2024 y tendrá que aplicarse de forma progresiva hasta el 26 de agosto de 2026. Pretende el establecimiento de un marco jurídico uniforme y armonizado vinculado al desarrollo, la «introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial» garantizando a la vez la libre circulación transfronteriza de mercancías y servicios basados en una IA fiable, segura y que respete los derechos fundamentales.

El Reglamento parte de la definición de lo que son los sistemas de IA a fin de diferenciarlos de los de *software* o de la programación tradicional. A tal efecto se indica que:

Una característica principal de los sistemas de IA es su capacidad de inferencia. Esta capacidad de inferencia se refiere al proceso de obtención de resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos y virtuales y a la capacidad de los sistemas de IA para deducir modelos o algoritmos, o ambos, a partir de información de entrada o datos (Reglamento UE 2024/1689: 12).

Este reglamento clasifica los sistemas de IA según su nivel de riesgo y prohíbe usos considerados peligrosos. En cuanto a la desinformación, partiendo de que la IA puede causar efectos negativos reales, se menciona la necesidad de prestar atención a los riesgos sistémicos derivados de posibles usos indebidos intencionados o de control en distintos

3. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (conocido como Reglamento de Inteligencia Artificial).

ámbitos, en concreto en el de la desinformación. El hecho de que los modelos puedan hacer copias de sí mismos, autorreplicarse o entrenarse con otros modelos puede dar lugar a sesgos, discriminación y riesgos, facilitando la desinformación (Reglamento UE 2024/1689: 110). Así mismo, las personas deben saber con claridad que están interactuando con una IA y los prestadores están en obligación de «detectar y mitigar los riesgos sistémicos que pueden surgir de la divulgación de contenidos que hayan sido generados o manipulados de manera artificial, en particular el riesgo de los efectos negativos reales o previsibles sobre los procesos democráticos, el discurso cívico y los procesos electorales, también a través de la desinformación» (Reglamento UE 2024/1689: 120).

Sabemos que la IA produce ingentes cantidades de contenidos sintéticos en ocasiones difíciles de diferenciar del generado por los seres humanos, lo que puede crear no solo riesgos de desinformación y de manipulación, sino también de fraude, suplantación de identidades y posibles engaños a los consumidores. Por tanto, los proveedores de contenido de IA (audio, imagen, vídeo, texto) deben «marcar» estos contenidos haciendo fácil la detección de que han sido creados o manipulados artificialmente (Reglamento UE 2024/1689: 133).

Un último apunte. Este Reglamento introduce el término de «ultrasuplantaciones» o «contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeje notablemente a personas, objetos, lugares, entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos». Sobre esto, el reglamento obliga a los responsables de estos sistemas de IA a etiquetar los resultados y a informar de que el contenido ha sido creado / manipulado artificialmente. Sin embargo, el reglamento también explica que esta obligación de transparencia ha de ser compatible con el derecho a la libertad de expresión y el derecho a la libertad de las artes y de las ciencias («en obra o programas manifiestamente creativos, satíricos, artísticos, de ficción o análogos») (Reglamento UE 2024/1689: 134).

3.1.3. Otros reglamentos y adopción de normas nacionales

Existen otros dos reglamentos de aplicación en la UE a lo largo de 2025. Por una parte, el conocido como Reglamento Europeo sobre la Libertad de los Medios de Comunicación (Reglamento (UE) 2024/1083 del Parlamento Europeo y del Consejo, de 11 de abril de 2024, por el que se establece un marco común para los servicios de medios de comunicación en el mercado interior y se modifica la Directiva 2010/13/UE). Por otra, el Reglamento (UE) 2024/900 del Parlamento Europeo y del Consejo, de 13 de marzo de 2024, sobre transparencia y segmentación en la publicidad política.

El Reglamento Europeo sobre la Libertad de los Medios de Comunicación, en pleno vigor desde agosto de 2025 (aunque con algunas disposiciones aplicables en distintos momentos entre 2024 y 2027) ofrece, fundamentalmente, normas para proteger el pluralismo y la independencia editorial de los medios de comunicación. En este último aspecto pone el acento en los medios de servicio público (art. 5.2), disponiendo que los Estados miembros garanticen su independencia editorial efectiva. De igual manera, se busca transparencia en la propiedad de los medios y evaluación en las concentraciones en el mercado de medios.

Es de interés en el ámbito de este trabajo cómo en numerosos apartados de este reglamento se menciona a las autoridades reguladoras, otorgándoles un importante papel como organismos independientes, tanto en la supervisión como en la aplicación de sus disposiciones. Así, se les insta a colaborar y a establecer herramientas para la cooperación normativa, el control de la publicidad estatal y «la creación de bases de datos de propiedad de los medios de comunicación a las autoridades u organismos reguladores nacionales, o a otras autoridades u organismos competentes». Incluso indica que los organismos reguladores deben contar con medios necesarios para cumplir sus funciones (plantilla, personal medios técnicos, recursos financieros, herramientas digitales) al igual que con competencias adecuadas que les permitan actuar en la correcta aplicación de la normativa sobre medios de comunicación, supervisando y respondiendo con agilidad las consultas (Reglamento (UE) 2024/1083: art. 4 y 36, 43).

En cuanto a la IA y la desinformación, el Reglamento Europeo sobre la Libertad de los Medios de Comunicación busca regular la relación con grandes plataformas en línea, protegiendo a los medios de comunicación de la eliminación injustificada de contenidos por parte de plataformas digitales de gran tamaño, debiendo notificar a los medios antes de proceder a ello:

[...] las plataformas en línea de alcance mundial actúan como puertas de entrada a los contenidos de los medios de comunicación, con modelos de negocio que tienden a eliminar la intermediación para el acceso a los servicios de medios de comunicación y a amplificar los contenidos polarizadores y la desinformación. Dichas plataformas son también proveedores esenciales de publicidad en línea, lo que ha desviado los recursos financieros del sector de los medios de comunicación, afectando a su sostenibilidad financiera y, como consecuencia, a la diversidad del contenido que se ofrece (Reglamento UE 2024/1083: art. 4).

La segunda de las normativas en vigencia desde 2025 es el Reglamento (UE) 2024/900 del Parlamento Europeo y del Consejo, de 13 de marzo de 2024, sobre transparencia y segmentación en la publicidad política. Entre otras cuestiones, este reglamento parte de la consideración de que la publicidad política puede ser un vector de desinformación, sobre

todo cuando no revela precisamente su naturaleza política o cuando está sujeta a técnicas de segmentación. De ahí que se establezcan medidas para contrarrestar la desinformación, propiciando la transparencia en los contenidos, el etiquetado e identificación clara de la publicidad política y la regulación del uso de técnicas de segmentación. También este reglamento busca frenar injerencias ilícitas de los denominados «patrocinadores» interesados en el fomento de la desinformación, especialmente en los procesos electorales. Así mismo, establece que los editores de publicidad política deben poner a disposición de las autoridades responsables de auditar o supervisar a los actores políticos la información requerida en los avisos de transparencia. E, incluso, este reglamento prevé la creación de un repositorio europeo de anuncios políticos en línea, gestionado por una autoridad designada, para almacenar y facilitar el acceso público a todos los anuncios políticos en línea publicados en la UE o dirigidos a sus ciudadanos o residentes.

En cuanto al derecho interno, algunos estados han aprobado normas nacionales. Es de especial interés en el contexto analizado en este capítulo la Online Safety Act 2023 (OSA), aprobada por el Parlamento de Reino Unido, que regula los servicios en línea que emplean IA, con especial énfasis en la seguridad de los usuarios y la eliminación de contenido ilícito y perjudicial (Reino Unido, 2023a).

Se trata de una ley avanzada en cuanto a los desafíos emergentes asociados a la IA y la desinformación y que otorga, precisamente, al mencionado OFCOM, importantes competencias y funciones. En este sentido, OFCOM está facultado para determinar normas de carácter técnico, realizar auditorías y emitir multas además de supervisar a las plataformas y establecer un comité asesor para abordar la difusión de información falsa en línea. Así mismo, impone nuevas obligaciones a las empresas tecnológicas para proteger a sus usuarios de contenidos que inciten al odio, provoquen violencia o difundan desinformación. Y también introduce sanciones penales por creación y difusión de imágenes manipuladas o falsas (*deepfakes*), especialmente si son de carácter íntimo, tipificando, además, nuevos delitos relacionados con el uso malicioso de contenido generado con IA (Reino Unido, 2023a; OFCOM, 2024e).

3.2. Inteligencia artificial, verificación e identificación de contenidos

La IA y las herramientas automatizadas se utilizan de diversas maneras para combatir la desinformación, tanto por parte de los reguladores del audiovisual europeos como de las plataformas en línea y los verificadores de hechos. A continuación, se presentan algunos de los usos más destacados:

- Desarrollo (o colaboración en el desarrollo) de herramientas y aplicaciones de IA para la gestión y presentación de datos. También para la detección de contenido y evaluación de la veracidad de los contenidos: las herramientas de IA se utilizan para detectar contenido ilegal o perjudicial en línea, incluyendo noticias falsas, discursos de odio y *deepfakes*.
- Uso de la IA para el análisis de algoritmos de las plataformas en línea a fin de identificar patrones que puedan estar promoviendo la difusión de desinformación.
- Apoyo a la moderación de contenidos: uso de la IA para detectar y eliminar contenido ilegal, como noticias falsas o discursos de odio. Además, se están considerando herramientas automatizadas, incluyendo la IA, para detectar de manera proactiva contenido perjudicial para niñas, niños y adolescentes.
- Verificación de hechos (*fact-checking*): empleo de la IA para identificar y desacreditar noticias falsas, *deepfakes* y otros tipos de desinformación. Desarrollo de herramientas en línea que permiten a los usuarios acceder a artículos de verificación de datos publicados por la red.
- Análisis de datos: empleo de la IA para analizar grandes conjuntos de datos con el fin de identificar tendencias y patrones en la difusión de desinformación. Esto ayuda a los reguladores y a otros actores a comprender mejor cómo se propaga la desinformación y a desarrollar estrategias más efectivas para combatirla.
- Monitoreo de campañas de desinformación: utilización de herramientas de IA para monitorear la cobertura periodística.
- Análisis del impacto de la IA: los reguladores están creando comités y grupos de trabajo para estudiar el impacto de la IA en los medios de comunicación y en la sociedad en general. AGCOM en Italia, por ejemplo, creó un comité sobre IA para analizar las implicaciones de los sistemas de IA.

Mencionaremos algunas iniciativas destacadas en este primer eje temático. Las primeras están referidas al desarrollo de herramientas como la creación del índice IDL (Information Disorder Level) que es una métrica basada en el juicio de humanos para evaluar el grado de veracidad de textos generados por máquinas. Esta herramienta ha sido desarrollada en el marco de una investigación de la Universidad de Bergen (Noruega) sobre la capacidad de los sistemas de IA generativa, como ChatGPT, para crear noticias falsas y contenidos inexactos (Dierickx, L., Lindén, CG., Opdahl, A.L., 2023). Lo incluimos en este trabajo porque se ha difundido su existencia y posibilidades de uso desde The European Digital Media Observatory, que reúne a *fact-checkers*, expertos en alfabetización mediática e investigadores para analizar la desinformación y que colabora con

organismos reguladores y plataformas en línea. La versión actual de la herramienta, que está en fase de desarrollo y que tiene como objetivo crear un conjunto de datos de entrenamiento, es gratuita y de acceso libre: <https://ohmybox.info/idl/index.php> Introduciendo un texto creado por IA el sistema calcula el IDL, ofreciendo un resultado entre 0 y 10 (siendo 10 la aproximación a la que refiere mayor veracidad).

Por otra parte, las Autoridades Estatales de Medios de Comunicación en Alemania (Die Medienanstalten) utilizan IA para detectar violaciones de la ley en Internet. Para ello emplean KIVI, una herramienta de IA desarrollada por la Autoridad Estatal de Medios de Renania del Norte-Westfalia junto con la empresa Condat AG de Berlín, que permite realizar búsquedas automatizadas de sitios web, plataformas de vídeo, redes sociales y servicios de mensajería —permite consultar hasta diez mil páginas cada día— (Die Medienanstalten, 2024a). KIVI hace posible detectar contenidos delictivos en internet (representaciones de violencia, incitación al odio o pornografía), clasificando posibles contenidos ilegales según grupos delictivos y elaborando una evaluación inicial de su gravedad. También en esta línea, el Conseil Supérieur de l’Audiovisuel de la Federación Valonia-Bruselas está optimizando sus métodos de investigación con herramientas de análisis automático basadas en IA (CSA, 2024a). Así, ha publicado los primeros resultados del seguimiento de contenido pornográfico en la red social X utilizando también KIVI, destacando la baja moderación de la plataforma respecto de estos contenidos (CSA, 2024b). Con el fin de adaptar la herramienta a las realidades de la Bélgica francófona, la CSA elaboró una lista de 250 palabras clave en francés e identificó 90 cuentas de supercontagadores para proporcionar a KIVI una base de aprendizaje que le permitiera iniciar el seguimiento.

Hemos señalado cómo la Ley de Servicios Digitales DSA establece un marco regulador que otorga a las autoridades y organismos reguladores competentes la facultad de acceder a datos y algoritmos de las plataformas digitales para garantizar un entorno en línea seguro y transparente. Así, los citados previamente de carácter nacional pueden tener competencias adicionales para acceder a datos y algoritmos de plataformas digitales.

3.3. Efectos de la inteligencia artificial, alfabetización mediática y otras estrategias de prevención

En líneas generales, los organismos reguladores del audiovisual trabajan en la idea de que la educación mediática es esencial para combatir la desinformación. Así, han venido desarrollando actividades conectadas con la IA / alfabetización mediática (audiovisual, digital, informacional...), enfocadas en promover el pensamiento crítico en comunicación, la

verificación de fuentes y el uso responsable de la tecnología. Hay iniciativas diversas que incluyen campañas de concientización, talleres, publicaciones y guías prácticas. Por otra parte, se ofrecen herramientas para detectar noticias falsas y contenido generado por IA, incluyendo la revisión de fuentes y la verificación de datos.

También se explora el potencial de la IA para mejorar la educación mediática a través de herramientas interactivas y programas de aprendizaje. Así mismo, se han venido desarrollando informes diversos sobre impacto, efectos y consecuencias de la IA, impulsando la investigación en la materia. Además, se está promoviendo la colaboración a nivel nacional e internacional para intercambiar buenas prácticas y coordinar acciones de alfabetización mediática contra la desinformación.

Por poner algunos ejemplos en la materia, en el caso del Consejo Audiovisual de Andalucía se ha desarrollado un informe sobre *Aplicaciones web de deepnude* con IA generativa en el que se constata cómo estas herramientas facilitan la creación de imágenes de personas desnudas que vulneran derechos de máxima protección y que justifican la actuación de los distintos poderes públicos para bloquear el acceso a estas páginas y exigir responsabilidades (2024a). También se ha consultado con expertos en materia de IA procedentes de distintos ámbitos y analizado el marco jurídico, preparando con toda esta documentación un informe unificado (2024b). En su edición del *Barómetro Audiovisual de Andalucía* mencionada, se ha incluido un apartado específico sobre IA (CAA, 2025: 149-156). En cuanto a investigación y transferencia del conocimiento, el CAA organizó en octubre de 2024 el I Seminario de Investigaciones y Estudios del Audiovisual sobre IA y desinformación del cual se desprende esta publicación. De igual manera, ha publicado un número de sus *Cuadernos del Audiovisual* (vol. 14, julio 2025) centrado en los avances, riesgos y desafíos de la IA en el sector audiovisual.

Por su parte, el Consell de l'Audiovisual de Catalunya ha enfatizado la importancia de la investigación para distinguir entre información y desinformación. También ha procedido a la premiación de trabajos que aplican la IA en la educación mediática en la XXI edición del programa eduCAC (CAC, 2024a). Cuentan, así mismo, con una Plataforma para la Educación Mediática (PEM) que organizó en 2024 la Jornada «Desinformación e Inteligencia Artificial» (CAC, 2024b). Y, en materia de desinformación se ha reclamado la veracidad en los contenidos difundidos, especialmente en situaciones de crisis (CAC, 2024c).

Desde el Consell del Audiovisual de la Comunitat Valenciana, en colaboración con la Universitat Jaume I de Castelló, se ha producido una serie de siete vídeos educativos que abordan temas de la alfabetización mediática, entre ellos el titulado «La desinformación y las *fake news*» (CACV, 2024a). También participó como entidad colaboradora de la guía *Introducción al uso de la IA en periodismo: Guía de referencias y modos*

de uso (Cano Orón, L.; López Merí, A.: 2024), que aborda posibilidades y desafíos éticos de la IA. De igual forma, el CACV ha alertado sobre las consecuencias de las noticias falsas y ha elaborado un informe sobre discursos de odio y desinformación en Internet (CACV, 2024b).

En Italia, la Autorità per le Garanzie nelle Comunicazioni ha creado un Comité sobre IA (AGCOM, 2024a). También ha promovido la educación mediática y digital entre menores a través de cursos de formación en las escuelas que siguen las directrices establecidas en la Resolución nº 177/24/CONTRA que abordan temas como la protección de la reputación *online*, el discurso de odio o la desinformación (AGCOM, 2024b). De igual forma, el regulador italiano ha intervenido para bloquear la emisión en YouTube y X del documental *Donbass, ayer hoy y mañana*, producido por *Russia Today*, por considerar que incitaba al odio y la desinformación (AGCOM, 2024c).

Por otra parte, la autoridad francesa ARCOM ha desarrollado recomendaciones con buenas prácticas dirigidas a las plataformas en línea para combatir la desinformación en periodos electorales (ARCOM, 2024a). En octubre de 2024, ARCOM organizó conferencias y mesas redondas sobre los usos e impactos de la IA en la comunicación audiovisual y digital (ARCOM, 2024b). En la 35ª Semana de prensa y medios se abordaron temas de actualidad, entre ellos, la IA (ARCOM, 2024c). En su *Informe sobre alfabetización mediática informacional 2022-2023*, ARCOM señala que evaluará la eficacia de las acciones emprendidas por los operadores de plataformas en línea para la lucha contra la manipulación de la información (ARCOM, 2023, p.14).

En Portugal, la Entidade Reguladora para a Comunicação Social ha participado en iniciativas para combatir la desinformación, como la jornada «*Fake news: Retos para la democracia y la sociedad*» celebrada en septiembre de 2024 (ERC, 2024a). También ha organizado campañas de alfabetización mediática, en especial de cara a las elecciones europeas (ERC, 2024b) y se pronunció sobre el contenido generado por IA incluido en un programa informativo (ERC, 2024c). ERC, en su *Deliberación ERC/2024/83 (OUT-NET)*, calificó de desinformación una publicación en la red X de encuestas políticas falsas sobre las elecciones legislativas de 2024 (ERC, 2024d). También ha definido diez ejes estratégicos para el mandato 2023-2028 que incluyen la lucha contra la desinformación (ERC, 2024e),

Las Autoridades Estatales de Medios de Comunicación en Alemania han desarrollado el informe sobre IA y enseñanza de la alfabetización mediática en el que se examina el impacto de la IA en el sector de los medios y se proponen medidas para fortalecer la alfabetización mediática (Die Medienanstalten, 2024d). En un taller sobre IA en 2024, la Comisión para la determinación de la concentración en el sector de los medios (KEK) debatió los efectos de la IA en este ámbito (Die Medienanstalten, 2024b). En su

Informe 2023, KEK muestra su preocupación por los efectos de la IA en los medios y la diversidad de opinión (Die Medienanstalten, 2024c).

Por su parte, en Reino Unido, OFCOM está trabajando para ayudar a los votantes jóvenes a identificar desinformación política. Para ello, realizó una campaña en diferentes redes sociales que ofrecía recomendaciones para identificar y combatir la desinformación (OFCOM, 2024a). OFCOM también ha publicado una carta abierta a los proveedores de servicios *online* explicando cuestiones referidas a IA generativa y chatbots en la mencionada *Online Safety Act 2023*, de la que es supervisor (OFCOM, 2024b). OFCOM ha realizado estudios sobre *deepfakes* y uso de la IA, tanto para su producción como para su detección (OFCOM, 2024c). Asimismo, el regulador británico publicó un informe donde explora las actitudes y comportamientos de los adultos del Reino Unido en relación con la desinformación e investiga cómo la gente percibe y responde a noticias falsas, incluyendo la frecuencia con la que las encuentran *online* en televisión o en prensa (OFCOM, 2024d).

En cuanto a las plataformas de organismos reguladores, el European Regulators Group for Audiovisual Media Services ha venido participando en iniciativas para combatir la desinformación durante las elecciones europeas. Así, en vísperas de las más recientes, inició junto con la Comisión Europea una campaña de comunicación conjunta para informar a los ciudadanos sobre los riesgos relacionados con la desinformación y la manipulación e injerencia de la información extranjera (ERGA, 2024a). También desde la ERGA se han impulsado campañas de alfabetización mediática en colaboración con operadores de televisión y plataformas digitales. En su 21.ª reunión plenaria se hizo balance de una de estas campañas que impulsó en colaboración con operadores de televisión y plataformas digitales (ERGA, 2024b).

La European Platform of Regulatory Authorities incluyó en su programa de trabajo para 2024 temas como la IA y la alfabetización mediática (EPRA, 2024a). Desde la EPRA también se han hecho eco de la publicación, por el Consejo de Europa, de las *Directrices sobre la implementación responsable de sistemas de IA en periodismo* (EPRA, 2023). De igual forma, han difundido el informe de noticias digitales 2024 del Instituto Reuters para el Estudio del Periodismo, que incluye datos sobre desinformación y destaca las redes sociales que generan mayor preocupación en cuanto a la dificultad para detectar noticias fiables (EPRA, 2024b).

Desde la Mediterranean Network of Regulatory Authorities se han difundido las directrices de la UNESCO para la gobernanza de las plataformas digitales y los desafíos y oportunidades que presenta la IA en el sector de los medios de comunicación y en materia de alfabetización mediática (MNRA, 2024). Y, por último, la Plataforma de Reguladores del Sector Audiovisual de Iberoamérica cuenta con un grupo específico de

trabajo en alfabetización mediática que ha elaborado un mapa de iniciativas en Iberoamérica. También desde la PRAI se han organizado jornadas en esta materia, así como sobre comunicación y democracia, para reflexionar sobre la desinformación, destacando la necesidad de normativas actualizadas y la importancia de la colaboración entre gobiernos, sociedad civil y las plataformas digitales (PRAI, 2024).

Conclusiones

Es evidente la preocupación generalizada de los gobiernos y de la ciudadanía acerca del peligro que supone la desinformación y cómo la IA está jugando un papel cada vez más preponderante en esta materia. Con el cambio hacia un nuevo paradigma en la comunicación que supone la irrupción de la IA, los organismos reguladores del audiovisual en Europa han venido trabajando en los importantes desafíos que la misma plantea. Son cada vez mayores las evidencias que demuestran que la IA puede ser utilizada tanto para detectar como para crear desinformación.

De ahí que los reglamentos que se han venido poniendo en marcha en la Unión Europea (fundamentalmente el *Reglamento de Servicios Digitales* y el *Reglamento de la UE sobre IA*) les hayan conferido funciones relevantes en la supervisión control de las obligaciones de los prestadores de plataformas en línea. Así, cumplen un papel de relevancia en la detección riesgos asociados a la divulgación de contenidos generados o manipulados con IA, en particular en lo referido a la desinformación.

En este ámbito, en este trabajo se han definido tres ejes temáticos que son comunes en las actuaciones y desarrollo de iniciativas de las autoridades regulatorias en Europa y de las plataformas que los asocian: IA, legislación y estrategias regulatorias; IA, verificación e identificación de contenidos, y efectos de la IA, alfabetización mediática y otras estrategias de prevención.

En cuanto a IA y legislación, en el marco de sus tradiciones jurídicas, las autoridades regulatorias han empezado a ejercer las funciones que les han sido encomendadas por los reglamentos estudiados (incluso, algunos países se han dotado de normas propias). Para ello han adaptados sus estructuras y también publicado informes diversos sobre la IA en el sector audiovisual, incluyendo sus implicaciones legales. Estos marcos jurídicos instan a los organismos reguladores a jugar un papel importante en el fomento de la colaboración entre verificadores de datos, investigadores y responsables políticos.

En cuanto al segundo eje, se ha puesto de manifiesto cómo la IA no solo genera desinformación, sino que las herramientas automatizadas están desempeñando un importante papel en la lucha contra la misma, la pérdida de confianza en las instituciones y la posible manipulación de

la opinión pública. Desde esta perspectiva, las autoridades reguladoras han desarrollado y promovido herramientas para evaluar la veracidad de los contenidos generados por IA, como las herramientas IDL o KIVI. También están considerando los riesgos éticos y sociales asociados con la IA, incluyendo los sesgos algorítmicos. De igual manera, han publicado indicadores para evaluar las medidas implementadas a objeto de mitigar la desinformación. Y, en definitiva, las autoridades reguladoras están poniendo en marcha iniciativas para investigar el impacto de la IA así como el desarrollo y compartición de herramientas y metodologías de IA y verificación de datos.

En cuanto al tercer eje, este trabajo muestra cómo los organismos reguladores europeos han comprendido la necesidad de desplegar esfuerzos para mejorar la alfabetización mediática en todos los grupos de edad y sectores de la sociedad para garantizar una ciudadanía informada en cuanto a IA y desinformación. En esta materia, coinciden en ofrecer guías y directrices sobre alfabetización mediática, así como conferencias y talleres. También en el desarrollo de recursos para la educación a distintos niveles sobre la desinformación, *deepfakes* y contenido generado por IA. Coinciden, además, en la importancia de evaluar críticamente los resultados de la IA y combinar su uso con otras estrategias, como la alfabetización mediática, la cooperación en acciones y la creación de códigos de buenas prácticas. En estos ámbitos se han mencionado distintas iniciativas que enfatizan el valor de compartir conocimientos para combatir la desinformación de manera eficiente.

Nota metodológica: la autora agradece la colaboración en la documentación de este trabajo del Área de Contenidos y del Área Jurídica del Consejo Audiovisual de Andalucía.

Referencias

- Autorità per le Garanzie nelle Comunicazioni (2024a, 24 enero). *Resolución 11/24/CONS de 24 de enero. Creación de un Comité de inteligencia artificial en las áreas de competencia de la Autoridad*. <https://bit.ly/3EbG1eU>.
- Autorità per le Garanzie nelle Comunicazioni (2024b, 5 de junio). *Resolución. 177/24/C. Aprobación de directrices en materia de cursos de formación en ciudadanía digital*. <https://bit.ly/4lk8WhB>.
- Autorità per le Garanzie nelle Comunicazioni (2024c, 13 junio). *Agcom interviene para bloquear contenidos procedentes de Russia Today difundidos por Youtube y X*. <https://bit.ly/4iVBgoJ>.
- ARCOM_ Autorité de régulation de la Communication Audiovisuelle et Numérique (2024a, 6 marzo). *Recomendaciones para luchar contra la manipulación de la información en las plataformas en línea con vistas a las elecciones al Parlamento Europeo del 6 al 9 de junio de 2024*. <https://bit.ly/4lrZBV9>.
- ARCOM_ Autorité de régulation de la Communication Audiovisuelle et Numérique (2024b, 14 octubre). *Arcom reúne a expertos sobre los impactos de la inteligencia artificial en los campos de la creación y la información*. <https://bit.ly/4i4W9N0>.
- ARCOM_ Autorité de régulation de la Communication Audiovisuelle et Numérique (2024b, 14 marzo). *Semana de Prensa y Medios en las Escuelas: La acción de Arcom en educación mediática, información y ciudadanía digital*. https://lc.cx/T_CLIP.
- ARCOM_ Autorité de régulation de la Communication Audiovisuelle et Numérique (2023, 7 diciembre). *Informe sobre alfabetización mediática e informacional (2022-2023, p.17)*. https://lc.cx/wEr_OT.
- Cano Orón, L., y López Merí, A. (2024). *Introducción al uso de la IA en periodismo: Guía de referencias y modos de uso*. *Publicacions De La Universitat De València*. <https://lc.cx/ioNf9P>.
- Comisión Europea (2018): *Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. La lucha contra la desinformación en línea: un enfoque europeo*. <https://lc.cx/376Bw2>.
- Consejo Audiovisual de Andalucía (2024a). *Aplicaciones web de deepnude con IA generativa*. <https://lc.cx/fUsyvN>.
- Consejo Audiovisual de Andalucía (2024b). *Informe del CAA sobre IA*. <https://lc.cx/fUsyvN>.
- Consejo Audiovisual de Andalucía (2025). *Barómetro del Audiovisual de Andalucía*. <https://lc.cx/kBQU3M>.
- Consell de l'Audiovisual de Catalunya (2024a, 29 mayo). *El CAC premia 15 trabajos que fomentan la educación en comunicación audiovisual en las escuelas*. <https://lc.cx/XE0ep3>.
- Consell de l'Audiovisual de Catalunya (2024b, 24 octubre). *El presidente del CAC pide que se garantice la veracidad y la trazabilidad de los contenidos generados por inteligencia artificial*. <https://kutt.it/SYQpVg>.

- Consell de l'Audiovisual de Catalunya (2024c, 8 noviembre). *El CAC, a raíz de la cobertura de la DANA, pide la comprobación de la veracidad de la información así como evitar la espectacularización*. <https://kutt.it/ADzLTP>.
- Consell del Audiovisual de la Comunitat Valenciana (2024a). *Alfabetización digital*. <https://cdacv.es/es/alfabetizacion-digital/>.
- Consell del Audiovisual de la Comunitat Valenciana (2024b). *Els discursos d'odi: Llibertat d'expressió, xarxes socials i esfera pública*. <https://kutt.it/Nszmxs>.
- Conseil Supérieur de l'Audiovisuel - Fédération Wallonie-Bruxelles (2024). *Les contenus pornographiques circulent quasi librement sur X/Twitter*. <https://kutt.it/kb5Y2d>.
- Conseil Supérieur de l'Audiovisuel - Fédération Wallonie-Bruxelles (2024a, 10 octubre). *La IA al servicio de la investigación de CSA*. <https://kutt.it/3665Vt>.
- Conseil Supérieur de l'Audiovisuel - Fédération Wallonie-Bruxelles (2024b, 5 febrero). *El contenido pornográfico circula casi libremente en X/Twitter*. <https://kutt.it/xMxbnE>.
- Die Medienanstalten (2024a, 22 febrero). *La aplicación efectiva de la ley puede fortalecerse a través de la DDG (Ley de Servicios Digitales)*. <https://kutt.it/zumatG>.
- Die Medienanstalten (2024b, 10 junio). *Taller de Inteligencia artificial de KEK. Expertos debaten sobre el impacto de la IA en los medios*. <https://kutt.it/VwY3aG>.
- Die Medienanstalten (2024c, 27 mayo). *KEK publica su 25º informe anual. La KEK está preocupada por los peligros de la diversidad que plantea la IA*. <https://kutt.it/9EZCCV>.
- Die Medienanstalten (2024d, abril). *Informe sobre la IA y la alfabetización mediática*. <https://kutt.it/gXdU5G>.
- Dierickx, L., Lindén, C. G. y Opdahl, A. L. (2023). The Information Disorder Level (IDL) Index: A Human-Based Metric to Assess the Factuality of Machine-Generated Content. En: Ceolin, D., Caselli, T. y Tulin, M. (eds.): *Disinformation in Open Online Media*. MISDOOM 2023. Lecture Notes in Computer Science, vol. 14397. Springer. <https://kutt.it/5LxRCy>.
- Entidade Reguladora para a Comunicação Social (2024a, 2 octubre). *ERC en el evento «Noticias falsas: retos para la democracia y la sociedad»* <https://kutt.it/mKDWhd>.
- Entidade Reguladora para a Comunicação Social. (2024b, 2 mayo). *ERC lanza una campaña de concienciación sobre los riesgos de la desinformación en periodo electoral*. <https://kutt.it/myfmFZ>.
- Entidade Reguladora para a Comunicação Social (2024c, 10 abril). *Decisión ERC/2024/178 (OUT-TV) Participación contra CNN Portugal por presentar un elemento basado en Inteligencia Artificial sobre los debates de los candidatos a las Elecciones Legislativas, el «Pulsómetro», en un programa informativo*. <https://kutt.it/sF4vZw>.
- Entidade Reguladora para a Comunicação Social (2024d, 5 marzo). *ERC publica recursos educativos sobre encuestas políticas*. <https://kutt.it/9Y47R5>.
- Entidade Reguladora para a Comunicação Social (2024e, 7 noviembre). *ERC presenta los ejes estratégicos para el periodo 2023-2028*. <https://kutt.it/haDbnB>.
- European Platform of Regulatory Authorities (2023, 14 diciembre). *El Consejo de Europa, publica directrices sobre la implementación responsable de sistemas de IA en el periodismo*. <https://kutt.it/8VLPbT>.

- European Platform of Regulatory Authorities (2024a, 12 febrero). La EPRA adopta el programa de trabajo para 2024. <https://kutt.it/RQqsce>.
- European Platform of Regulatory Authorities (2024b, 8 julio). La EPRA adopta el programa de trabajo para 2024. <https://kutt.it/a3b4F5>.
- European Regulators Group for Audiovisual Media Services (2024a, 8 mayo). La Comisión impulsa la concienciación sobre los riesgos de la desinformación y la manipulación de la información. <https://kutt.it/8VKpUM>.
- European Regulators Group for Audiovisual Media Services (2024b, 5 julio). ERC participa en la 21ª reunión plenaria del Grupo Europeo de Reguladores Audiovisuales. <https://erga-online.eu/?p=1514>.
- Mediterranean Network of Regulatory Authorities (2024, 26 noviembre). 23ª Asamblea Plenaria, 8 y 9 de octubre de 2024, Ereván, Armenia. <https://kutt.it/EDKg2C>.
- OFCOM (2024a, 19 junio). Ayudando a desmentir la desinformación en torno a las elecciones generales. <https://kutt.it/hB8qtC>.
- OFCOM (2024b, 8 noviembre). Carta abierta a los proveedores de servicios en línea del Reino Unido sobre la IA generativa y los chatbots. <https://kutt.it/DFcWV6>.
- OFCOM (2024c, 23 julio). Una inmersión en las deepfakes que degradan, defraudan y desinforman. <https://kutt.it/xGRGRX>.
- OFCOM (2024d, 27 noviembre). Comprender la desinformación: un análisis del comportamiento y las actitudes de los adultos del Reino Unido. <https://kutt.it/EM6tRu>.
- OFCOM (2024e, 5 agosto). Abordar el contenido en línea que incita al odio, provoca la violencia y difunde desinformación. <https://kutt.it/Qhxpev>.
- Plataforma de Reguladores del Sector Audiovisual de Iberoamérica (2024, 8 julio). Nueva Jornada abierta PRAI «Comunicación y democracia: ¿Quién regula la verdad?». <https://kutt.it/rxCL2W>.
- Gobierno de Reino Unido (2023). Online Safety Act 2023. <https://kutt.it/f6FSrr>.
- Unión Europea (2022). Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales). <https://kutt.it/dsnMgy>.
- Unión Europea (2024). Reglamento (UE) 2024/1083 del Parlamento Europeo y del Consejo, de 11 de abril de 2024, por el que se establece un marco común para los servicios de medios de comunicación en el mercado interior y se modifica la Directiva 2010/13/UE (Reglamento Europeo sobre la Libertad de los Medios de Comunicación). <https://kutt.it/82BeCQ>.
- Unión Europea (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). <https://kutt.it/M9DxaS>.

Inteligencia artificial contra la desinformación: análisis internacionales sobre plataformas de *fact-checking* iberoamericanas y europeas

María Sánchez González

Universidad de Málaga
m.sanchezgonzalez@uma.es

Hada Sánchez Gonzales

Universidad de Sevilla
misago@us.es

1 | Introducción

El *hype* reciente de la inteligencia artificial, en especial de la IA generativa y de herramientas orientadas a la creación de contenidos en formatos diversos (texto, imágenes, música, código...) ha traído consigo un ingente debate que atañe a empresas, gobiernos, profesionales y ciudadanía. De ahí la necesidad de establecer medidas y pautas de regulación como señalan otros autores (Bontridder & Pouillet, 2021).

Al margen de sus riesgos y peligros, es innegable que estas tecnologías disruptivas están siendo una verdadera palanca para la innovación

y el emprendimiento en diversos sectores (Shepherd & Majchrzak, 2022), incluyendo la comunicación.

En el ámbito concreto del periodismo y de la verificación de información, la IA se ha convertido en herramienta clave para luchar contra la llamada infodemia, localizando «la información falsa que llega a las redacciones para conseguir transparencia y veracidad de la información» (Sánchez, Sánchez & Martínez, 2022). Hablamos del denominado «*fact-checking* automatizado» (García, 2022), que está marcando una nueva era en el periodismo (Porlezza, 2023; Sánchez, Sánchez & Alonso, 2024). Este fenómeno ha cobrado protagonismo en los últimos años, como práctica que posibilita la comprobación automática de las noticias, según categorías utilizadas (falso, verdadero, etcétera) y distintos enfoques y métodos (Parikh y Atrey, 2018) relativos a análisis de atributos lingüísticos y comportamiento de los usuarios en redes sociales (*ibidem*).

Los profesionales de medios de comunicación con secciones o de proyectos de *fact-checking* y plataformas independientes específicamente orientadas a la verificación de información vienen por tanto empleando IA por iniciativa propia o como parte de políticas empresariales. Al tiempo, determinadas empresas invierten recursos en desarrollar sus propias herramientas basadas en IA, ya sea para uso interno de sus verificadores y periodistas, ya sea para abrirlas a la ciudadanía, e incluso hay entidades que diversifican su actividad con recursos e iniciativas orientadas precisamente a la alfabetización y capacitación de diversos colectivos y público (Sánchez, Sánchez y Alonso, 2024).

En este contexto, el presente capítulo recoge el planteamiento y los principales resultados y conclusiones de dos análisis realizados a nivel internacional, sobre plataformas de verificación del ámbito iberoamericano en 2022 y de Europa en 2024, respectivamente. Primero se abordaron los verificadores del ámbito americano que formaban parte de la Red IFCN (International Fact-Checking Network), miembros de #CoronavirusFacts Alliance, siendo un total de 16 verificadores entre los que se situaban 3 españoles, Maldita, Newtral y EFE Verifica. Y más adelante se estudiaron verificadores del ámbito europeo, concretamente los miembros de la Red EFCSN (European Fact-checking Standards Network), siendo un total de 10, entre los que estaban también incluidos tres españoles: Maldita, Newtral y Verificat en este caso. Ambos son censos de verificadores de referencia, cuya actividad se ajusta a determinados estándares profesionales (Sánchez, Sánchez y Martínez, 2022; Sánchez, Sánchez y Alonso, 2024). Y ambos análisis se abordaron con un planteamiento similar, descrito más adelante, lo que facilita disponer de una evolución en cuanto a los aspectos abordados de los casos españoles, al estar incluidas las mismas plataformas en los dos.

Tales aspectos fueron los siguientes (objetivos de análisis):

- Conocer el empleo que realizan los verificadores de referencia iberoamericanos y europeos de la inteligencia artificial en procesos vinculados con *fact-checking* o su papel activo contribuyendo al desarrollo de herramientas propias, entre otros aspectos.
- Determinar la percepción de tecnólogos y de responsables de estas plataformas de verificación sobre la inteligencia artificial, sus riesgos y sus posibilidades contra la desinformación; así como si su empleo responde a determinada estrategia organizacional, para unos objetivos concretos y al servicio de la innovación, o se realiza más bien de forma táctica, y la visión sobre IA y verificadores a medio/largo plazo.
- Identificar iniciativas de capacitación y/o alfabetización sobre IA y verificación desarrolladas por los *fact-checkers*, ya sean actividades, ya sean recursos; presenciales, híbridas o en red; dirigidas a ciudadanía o a periodistas u otros colectivos profesionales determinados.

Estos trabajos han sido realizados en el seno de varios proyectos de I+D+i en que vienen participando sus autoras. En concreto, un proyecto de investigación financiado por el gobierno de España (Ministerio de Ciencia, Innovación y Universidades de España), «El Impacto de la Desinformación en el Periodismo: Contenidos, Rutinas Profesionales y Audiencias» (PID2019-108956RB-I00) y su continuación «Aplicaciones periodísticas de la IA para reducir la desinformación: tendencias, usos y percepciones de profesionales y audiencias (DESINFOPERIA)» (PID 2023-147486OB-I00); el proyecto «Impacto de la desinformación en Andalucía: análisis transversal de las audiencias y las rutinas y agendas periodísticas. Desinfoand» (PAIDI 2020, ProyExcel_00143), de la Junta de Andalucía, España; y «genIA: Biblioteca de herramientas de inteligencia artificial para la creación de contenidos en medios de comunicación», del II Plan Propio de Investigación, Transferencia y Divulgación Científica de la Universidad de Málaga (2023 y 2024).

2 | Metodología

Las investigaciones cuyos resultados se presentan en este capítulo abordan el uso de la inteligencia artificial al servicio del *fact-checking* por verificadores hispanos y europeos y, como se adelantó, desde una doble perspectiva: la percepción de esta tecnología y de su utilidad para sus promotores, por un lado; y la actividad de estos verificadores referente a la IA, en cuanto a su empleo como herramienta para verificar, al desarrollo tecnológico de herramientas propias y a la alfabetización/ formación vinculadas a esta tecnología.

Ambos estudios, el del caso hispano y del caso europeo, se han realizado de forma similar, combinando distintas técnicas y a lo largo de varias fases.

En una fase inicial, con vistas a sustentar tanto el marco teórico como el diseño metodológico de ambos análisis, se realizó una revisión documental, complementada con la visión de expertos en tecnología e innovación periodística a través de entrevistas sobre IA en verificación y antecedentes de análisis sobre su uso por medios. Para las entrevistas se localizaron 8 perfiles (Rocío Celeste, experta en IA; Ana Valdivia, experta en datos; José Carlos Sánchez, de Prodigioso Volcán; Miriam Hernanz, experta en nuevas narrativas, antes en LabRTVE; Eirini Chatzikoumi, investigadora periodismo e IA; Idoia Salazar, periodista y presidenta de Odisela; Javier Cantón, experto en datos y verificación, entonces en Verifica RTVE; y David Fernández, de Maldita). Se realizaron entre febrero y abril de 2021, en un periodo en el que apenas había precedentes, por videoconferencia. Entre otras cuestiones se recogió su percepción acerca de las ventajas que proporciona la implantación de la IA en las redacciones, las limitaciones que surgen en su uso, así como el papel que juegan los periodistas en el manejo de este tipo de tecnología.

En una fase posterior, con planteamiento similar para ambos censos de verificadores y como núcleo central de la investigación, se procedió a identificar sus sitios web y a localizar, analizar y catalogar sobre criterios tecnológicos y funcionales proyectos innovadores con/sobre IA, empleando un formulario online de recogida de datos. Mientras que en el caso iberoamericano esta fase se desarrolló entre marzo y mayo de 2022, en el europeo se efectuó más tarde, a inicios de 2024.

Lo primero fue determinar qué verificadores habían desarrollado herramientas de IA, y en caso afirmativo se procedió a su análisis pormenorizado, recogiendo datos en cuanto a sus rasgos y funcionalidad como su denominación; tipología (*microsite*, extensión de navegador, aplicación online, bot, aplicación móvil, servicio u otras); orientación/ utilidad respecto a la labor periodística (generación automática de contenido, procedimientos de verificación, gestión y depurado de ingentes bases de datos, envío de respuestas/contenidos automatizados a las audiencias, u otras); uso en procesos de verificación (verificación de fuentes, verificación en el procesamiento e interpretación de textos, verificación de frases previamente publicadas, verificación del audio en texto, u otras); año de creación y disponibilidad en el momento del análisis (sí/no), y apertura o no a las audiencias (Sí/No). Para cada herramienta recogimos también información adicional (contexto de creación, vinculación a iniciativas o proyectos de innovación tecnológica o editorial del medio, colaboraciones externas en su desarrollo y tipo de entidades colaboradoras, financiación, etcétera).

En cuanto a la puesta en práctica de iniciativas de alfabetización (actividades y recursos) sobre IA y verificación, en los verificadores en que

localizamos algunas se fueron radiografiando en función de sus parámetros básicos (actividad/ recursos; modalidad; enfoque u objetivo; destinatarios; etcétera).

Como se anticipó en la introducción, se analizaron, por un lado, el total de verificadores de la Red IFCN miembros de #CoronavirusFacts Alliance (Sánchez, Sánchez & Martínez, 2022), incluidos verificadores españoles de la red (n=16). Estos incluyeron los españoles Maldita, Newtral y EFE Verifica (n=3, con entrevistas a Clara Jiménez, Joaquín Ortega/Irene Larraz y Desireé García respectivamente). Además de Agencia Lupa; Agencia Ocote-Fáctica (entrevista a Alejandra Gutiérrez); Animal Político-El Sabueso (E: Tania Montalvo); Aos Fatos; Bolivia Verifica; Chequeado (E: Laura Zommer); Colombia Check (E: Jeanfreddy Gutiérrez); Ecuador Chequea (E: Erika Astudillo); Estadão Verifica; La Silla Vacía-Detector de mentiras (E: María Echeverry); Observador; Salud con Lupa-Comprueba (E: Fabiola Torres); y La República-Verificador (E: Irene Ignacio).

Por otro lado, se investigó el total de la Red EFCSN (European Fact-checking Standards Network) (n=10) (Sánchez, Sánchez & Alonso, 2024). Un censo compuesto por los verificadores españoles Maldita, Newtral y Verificat (n=3, con entrevista a Pablo Hernández en el primer caso), y por los europeos AFP F actual (E:Natalia Sanguino); Facta.news; Demagog Association (E: Aleksy Szymkiewicz); Pagella Política; Correctiv (E: Caroline Lindkamp); Teyit; y TjekDet (E: Thomas Hedin).

- Verificadores Red IFCN de Iberoamérica
 - Agencia Lupa
 - Agencia Ocote-Fáctica (E: Alejandra Gutiérrez)
 - Animal Político-El Sabueso (E: Tania Montalvo)
 - Aos Fatos
 - Bolivia Verifica
 - Chequeado (E: Laura Zommer)
 - Colombia Check (E: Jeanfreddy Gutiérrez)
 - Ecuador Chequea (E: Erika Astudillo)
 - EFE Verifica (E: Desireé García)
 - Estadão Verifica
 - La Silla Vacía-Detector de mentiras (E: María Echeverry)
- Verificadores red EFCSN de Europa
 - Maldita-Maldito Buló (E: Clara Jiménez)
 - Newtral (E: Joaquín Ortega/Irene Larraz)
 - Observador
 - Salud con Lupa-Comprueba (E: Fabiola Torres)
 - La República-Verificador (E: Irene Ignacio)
 - Maldita (E: Pablo Hernández)
 - AFP Factual (E:Natalia Sanguino)
 - Facta.news
 - Demagog Association (E: Aleksy Szymkiewicz)

- Pagella Política
- Correctiv (E: Caroline Lindekamp)
- Teyit
- Newtral
- TjekDet (E: Thomas Hedin)
- Verificat

Como fase final se recabó la visión de los responsables de los verificadores seleccionados mediante las entrevistas ya citadas. Para los iberoamericanos usamos datos extraídos de entrevistas realizadas para un análisis anterior a algunos de sus responsables, que fueron actualizadas también mediante contacto con éstos entre abril y mayo de 2022. En el caso de los europeos se contactó con ellos online a inicios-mediados de 2024.

Como resultado obtuvimos un censo de 11 entrevistas (de los 16) de la IFCN y 5 entrevistas (de los 10) de la EFCNS (2024).

En estas entrevistas se buscó información adicional sobre los proyectos; datos sobre el contexto para la innovación tecnológica y el uso de IA en los verificadores analizados (existencia o no de secciones de I+D+i, perfiles, percepción de potencial de IA, participación en programas o colaboraciones externas...). También se indagó sobre la percepción que sobre la IA tenían los responsables de los verificadores, tanto en general (amenazas y oportunidades) como con relación a sus proyectos (fortalezas, debilidades). Y se recogieron sus planes de desarrollar iniciativas propias basadas en esta y otras tecnologías disruptivas a corto y medio plazo.

3 | Principales resultados

A continuación, se expone un compendio de los principales hallazgos de nuestros análisis, realizando una comparativa de los casos iberoamericano, europeo y español en los aspectos clave y aportando, para finalizar, un conjunto de casos significativos de innovación e intraemprendimiento en materia de IA aplicada a la verificación.

3.1. Aspectos en común entre los casos hispano y europeo

Se observa, en primer lugar, una coincidencia significativa en cuanto a la percepción positiva de la inteligencia artificial como herramienta de verificación, como complemento a la labor profesional que tienen los responsables de los dos conjuntos de verificadores analizados. Pero por encima de esta percepción favorable existe una elevada preocupación

por los riesgos asociados a estas tecnologías emergentes. En concreto, se manifiesta un desasosiego sobre el papel de la IA como potencial facilitador de *deep fakes* y, en general, de proliferación de desinformación en redes.

Es significativo que entre los perfiles profesionales de los equipos de las plataformas de *fact-checking*, aunque siga predominando el de periodista, aparezcan también algunos tecnológicos (programadores/desarrolladores informáticos o analistas). Una multidisciplinariedad necesaria ante proyectos complejos tecnológicamente. En ocasiones y según su dificultad se detectan también colaboraciones de expertos en determinadas materias o humanistas especializados en PNL.

En la práctica totalidad de entrevistas también se detectó que el uso de la IA no suele obedecer a una estrategia organizacional fijada de antemano, sino que más bien este comenzó de forma táctica, a modo de ensayo-error y como respuesta a necesidades específicas. Este uso estaba principalmente limitado a herramientas de creación de empresas externas y no de desarrollo propio, y se realizaba para la verificación de imágenes y de vídeo fundamentalmente. Se empleaba sobre todo en fases de investigación y verificación de fuentes de información como parte de su metodología. Procesos que han sido abordados por otros autores sobre estrategias comunes de metodología de verificación (Sánchez, Sánchez y Martos, 2024:6): «a) identification and selection of content-criteria and technique; b) research and verification; c) writing and publishing fact-checking results; and d) adhering to the transparency and correction policy». La tendencia de los *fact-checkers*, además, era emplear más de una herramienta para un mismo proceso de verificación.

En cuanto a la tipología de estas, hallamos un uso predominante de bots y chatbots, que coinciden con las herramientas de desarrollo propio. La creación de estas herramientas se da en casos específicos, principalmente en verificadores orientados a la filosofía de *startups* e innovación tecnológica. Estos, además, cuentan con equipos multidisciplinarios propios o recurren a apoyos externos para su puesta en marcha.

En nuestro análisis sobre la red de la IFCN (2022), localizamos 14 herramientas de IA propias, desarrolladas por 8 *fact-checkers* distintos, lo que supone que la mitad de los iberoamericanos analizados (16 en total) aportó sus propias creaciones al campo de la verificación asistido con IA. Los detalles sobre estas herramientas se verán más adelante

En contraste, en el panorama europeo, entre los *fact-checkers* de la red de la EFCNS (2024) se detectaron 10 herramientas basadas en IA desarrolladas por 7 de los 10 proyectos analizados. Esta mayor proporción en Europa podría atribuirse quizás a la diferencia temporal y la consecuente evolución de la IA en los últimos años. En el contexto europeo el porcentaje es mayor y la inmensa mayoría de los verificadores ha puesto en marcha sus propias herramientas para combatir los bulos, reflejando

así una tendencia creciente hacia la integración de la IA en las rutinas de *fact-checking*.

Esta comparación destaca el progresivo uso de la IA en la verificación de hechos en Europa con un avance e implementación de estas herramientas que proporcionan, por un lado, mejoras en la eficiencia del proceso de verificación y, por otro, la adaptación del sector a los cambios devenidos por la desinformación y la tecnología.

3.2. Divergencias o singularidades entre los verificadores

Los resultados de los análisis del caso iberoamericano y europeo divergen, más que por ámbito geográfico, según la tipología de verificadores. Las principales diferencias se observan sobre todo entre los micromedios o verificadores emergentes y los verificadores más consolidados. Estos últimos, como sería el caso de Chequeado en Argentina o Maldita en España, entre otros, generalmente disponen de mayores recursos, equipos más estables y diversos y operan bajo una cultura de la innovación y estrategias más asentadas.

Esta diferenciación es significativa porque evidencia la evolución de la verificación en el uso de inteligencia artificial en los últimos años. Muchos de ellos han pasado de ser meros usuarios a creadores o cocreadores de herramientas basadas en IA o impulsores de proyectos de alfabetización o capacitación como decíamos.

El análisis revela escasas diferencias entre el territorio iberoamericano y europeo en cuanto al uso de esta tecnología disruptiva y otros aspectos analizados, mostrando únicamente ciertos matices en cada caso. Por ejemplo, entre los verificadores iberoamericanos parece existir mayor orientación hacia lo social y la cultura abierta y casi todas las herramientas de desarrollo propio (un 78,6% de las localizadas) están disponibles para las audiencias. En el caso europeo, sin embargo, si bien están más concienciados sobre la importancia de instaurar algún tipo de política o normativa de uso de la IA en su labor de verificación, niegan que la IA sea una herramienta para que la ciudadanía pueda por sí misma verificar la información. Asimismo, orientan las herramientas de desarrollo propio o a sus profesionales de la información y verificación. Por lo tanto, la mayoría no están disponibles para uso público.

En Europa la creación y desarrollo propio de herramientas de IA se realiza, por lo general, mediante colaboraciones, ya sea con entidades públicas, ya sea con organizaciones privadas. En Iberoamérica, la mayoría de estas surgen en el seno de procesos de verificación y, en ciertos casos, en el marco de la IFCN, a través de proyectos colaborativos entre plataformas de la red, o mediante colaboraciones con desarrolladores externos.

3.3. Casos significativos y prácticas innovadoras

Concluimos la exposición de resultados apuntando tendencias significativas de IA para *fact-checking*, tanto de desarrollo tecnológico como de alfabetización, con marcadas diferencias territoriales.

En primer lugar, en el ámbito de verificación de países iberoamericanos destacan casos relevantes de innovación tecnológica con IA. Entre los *fact-checkers* veteranos, Chequeado lidera el desarrollo de herramientas propias basadas en inteligencia artificial. Un ejemplo es el proyecto Chequeabot, un bot pionero y versátil que ha ido evolucionando desde el 2016, que escanea medios de comunicación y escucha discursos parlamentarios para identificar aquellas frases que pueden ser verificables. Esta funcionalidad libera a los periodistas de tareas repetitivas para que puedan dedicar su tiempo «al análisis con valor agregado y no en las tareas más repetitivas que son relegadas al robot» (Sánchez, Sánchez y Martínez, 2022). A ello le sumamos que el bot ha incluido la verificación de *podcasts* para ampliar su alcance de verificación.

Los programas de innovación tecnológica y/o impulso a la alfabetización de las grandes empresas tecnológicas son un apoyo para que las microplataformas iberoamericanas generen sus propias IAs. Muchos programas de apoyo y *grants* combinan ambos objetivos (desarrollo tecnológico y alfabetización sobre IA), de ahí que varias herramientas iberoamericanas desarrolladas por verificadores de este territorio lleven también integradas «funciones» de alfabetización. Así nacieron herramientas como Chatbot Fatima, ¡de AOS Fatos y Projeto Lupe! de Agência Lupa, ambas iniciativas de alfabetización mediática que surgen en Facebook y en Brasil. Fatima pone a disposición del usuario información y sugerencias sobre cómo analizar noticias o cómo encontrar datos y fuentes fiables; mientras que el Projeto Lupe! de Agência Lupa, creada con el objetivo de informar a los usuarios sobre el proceso electoral de 2018, incluía vídeos y piezas educativas sobre los comicios para alfabetizar a la población.

En segundo lugar, en Europa, en lo que a desarrollo tecnológico con IA se refiere predominan dos vertientes: los *bots* y aplicaciones en línea. En la primera de ellas destaca el chatbot de Correctiv que sigue similar trayectoria que el de WhatsApp de Maldita en España. Este último, además, es reconocido como uno de los pioneros en proporcionar información accesible y verificada. Respecto al segundo caso, sobresale AI4TRUST, de Demagog Association, una «*platform for automating information verification*» que combina la IA con la experiencia humana o Vision creada por TjekDet, una «*testing tool which can locate claims in longer text*».

En materia de alfabetización se han implementado diversas iniciativas mediante guías y talleres de alfabetización sobre IA. En el primer caso, el verificador, TjekDet, proporciona guías de alfabetización con consejos

útiles, además de realizar talleres presenciales de alfabetización, no solo a estudiantes y profesionales, sino también a docentes. Esta última práctica, la de los talleres, también la realiza el verificador Correctiv, que en su caso se centra en colectivos vulnerables (tercera edad), reconociendo de esta forma la importancia de estos perfiles. En definitiva, se trata de iniciativas que destacan por su enfoque integral entre el desarrollo de sistemas de IA, la educación y concienciación a la población en la era de la desinformación.

Por último, en España, cuyos verificadores fueron analizados tanto en la oleada de 2022 como parte de la IFCN como en la de 2024 como parte de la EFCNS, recordemos, es posible destacar la evolución del uso de herramientas de IA, no solo orientadas a procesos de comprobación, sino de interpretación de textos. En 2024 son creadores de herramientas de desarrollo propio, así como disponen de protocolos y políticas de uso de inteligencia artificial. También en materia de desarrollo tecnológico con IA en España, podemos reseñar casos de verificadores que se han apropiado de la IA para solventar necesidades concretas, como sucede con Newtral en España. Esta plataforma ha centrado esfuerzos en crear herramientas tecnológicas que les permite detectar las frases verificables dentro del discurso político, a diferencia del resto de verificadores europeos que lo hacen mediante sus profesionales.

En cuanto a la alfabetización sobre IA por los verificadores españoles, cabe reseñar las diversas propuestas de Maldita. Esta labor se realiza sobre todo en su división Maldito Bulo, donde viene innovando en narrativas y formatos en redes sociales como Tiktok para concienciar y enganchar a los jóvenes (Sánchez, Sánchez y Martos, 2022b). Organiza sus propias formaciones desde hace tiempo. Además, su colaboración con universidades y otros agentes (organismos e instituciones públicas y privadas) es habitual, ya sea para crear contenidos digitales de formato innovador, como la *escape room* «La Fábrica de Bulos», con desarrollo tecnológico de la Universidad Politécnica de Madrid; ya sea mediante proyectos de investigación y transferencia. Un ejemplo de este último es el Iberian Digital Media Observatory (IBERFIER), coordinado desde la Universidad de Navarra y en el que Maldita colabora con otros verificadores españoles (EFE Verifica, Newtral y Verificat) y portugueses (Polígrafo y Prova dos Factos-Público).

Verificadores de surgimiento más reciente, como Verificat en Cataluña, son un ejemplo de *startups* que diversifican su actividad más allá del *fact-checking* y emplean la formación, con frecuencia dirigida a colectivos profesionales o universitarios futuros periodistas, como parte de su modelo de negocio. Impartir talleres y otros cursos les permite obtener recursos para financiarse o expandir su función social. Destacan, en su caso, iniciativas como la «Cápsula dentro del proyecto Desfake», que no solo se destina solo a estudiantes, sino también a colectivos vulnerables.

Conclusiones

Ante un panorama en el que también han surgido «pseudoverificadores» que utilizan estrategias similares a los verificadores legítimos para lograr la credibilidad, mientras difunden información engañosa con tecnologías recientes capaces de generar *deepfake* (Croitoru *et al.*, 2023), este análisis es relevante por varios motivos.

Para comenzar, se evidencia la relevancia de plataformas de *fact-checking* de referencia, integradas en redes internacionales y que desarrollan su actividad conforme a determinados principios de calidad y transparencia. Los casos reseñados, tanto en Iberoamérica como en Europa, suponen una muestra de intraemprendimiento e innovación, tanto tecnológica como periodística, en el ámbito de la verificación de noticias asistida con inteligencia artificial. De ahí la percepción positiva de sus responsables, que ven a la IA como una herramienta que apoya los procesos de verificación, aunque expresan su preocupación por los riesgos que supone la creación de *deepfakes* e incremento de desinformación. Con estructura de micromedios, en muchos casos, el uso y apropiación de tecnologías disruptivas como la inteligencia artificial generativa juega un papel estratégico, no solo para desarrollar su labor de *fact-checking* de forma más eficaz y eficiente, sino también para llegar más lejos y poner incluso la IA para verificar al servicio de profesionales y/o ciudadanía, con herramientas de desarrollo propio o capacitación sobre la materia.

El estudio comparado también revela algunos aspectos sobre la evolución y tendencias, diferencias respecto al ámbito geográfico y en innovación y alfabetización de los *fact-checkers* iberoamericanos y europeos de referencia.

En primer lugar, se comprueba la integración de perfiles profesionales en los equipos de verificación, aunque el perfil de periodista es el que predomina respecto al resto. También es perceptible la evolución del uso de la inteligencia artificial desde un enfoque táctico en respuesta a necesidades concretas de verificación con herramientas externas hacia un enfoque estratégico con el desarrollo de herramientas propias. Es último, sobre todo, en el ámbito europeo. Igualmente, existe una diferencia entre los verificadores consolidados y los emergentes respecto a la capacidad de innovación y recursos. Sin embargo, en ambos utilizan la IA para verificar sobre todo imágenes y videos.

En segundo lugar, con relación al ámbito geográfico es posible señalar a grandes rasgos dos matices que diferencian a los verificadores de Iberoamérica respecto a Europa. En el primer caso muestran una cultura abierta poniendo a disposición de la ciudadanía el uso de herramientas asistidas con IA. En el segundo caso, existe mayor conciencia reflexiva respecto a establecer políticas en el uso de la inteligencia artificial, aunque se limita el acceso de herramientas a los profesionales de la información y verificación.

En tercer lugar, entre los casos más significativos de innovación, destacan Chequeando en Iberoamérica con Chequeabot y TjekDet en Europa con guías y talleres de alfabetización. Respecto a los verificadores de España se distinguen, además del desarrollo propio de herramientas por los verificadores, por su enfoque en la alfabetización dirigido especialmente a colectivos vulnerables y jóvenes.

La presente investigación proporciona asimismo datos relevantes sobre el estado actual de la inteligencia artificial en los procesos de verificación, revelando oportunidades como desafíos a los que se enfrentan en su metodología con herramientas asistidas por la IA. En este sentido y como línea futura, sería interesante seguir investigando sobre su impacto a largo plazo. Si bien, autores en un estudio internacional han determinado estrategias y etapas comunes en el proceso metodológico (Sánchez, Sánchez y Martos, 2024), sería necesario saber el desarrollo gradual de su uso en cada fase. Así como el estudio paralelo de la evolución de los perfiles de *fact-checking* y, por consiguiente, la preparación de los estudiantes, futuros periodistas mediante un currículum formativo acorde con la demanda de las exigencias del mercado laboral.

Este estudio a pesar de ser internacional y comparativo también tiene algunas limitaciones que se centran en la muestra de verificadores de dos ámbitos geográficos: Iberoamérica y Europa, así como de dos redes internacionales como la IFCN y EFCSN. A esto le sumamos el factor del espacio temporal de la investigación realizada entre 2022 y 2024, respectivamente, pero que queda de base para próximas investigaciones en este ámbito de estudio por su utilidad.

La sinergia y colaboración de los verificadores con instituciones u organizaciones a nivel internacional resulta ser un factor clave para el desarrollo de futuras innovaciones (innovación abierta) y como líneas futuras de estudio, así como los desafíos éticos que plantea la transparencia de la información con el creciente uso de la IA. Esto podría afectar posiblemente a la calidad y efectividad de la verificación a largo plazo si no existe un uso responsable. Por lo tanto, es imprescindible avanzar en la investigación sobre la alfabetización de la IA y su efectividad en la ciudadanía.

Agradecimientos

Este trabajo forma parte de los proyectos «Desinforper (PID2019-108956RB-I00); Desinfoand» (PAIDI 2020, ProyExcel_00143) y «gen-IA: Biblioteca de herramientas de inteligencia artificial para la creación de contenidos en medios de comunicación» (II Plan Propio Universidad de Málaga, 2023).

Agradecimiento especial a todas/os las/los profesionales de verificadores y expertos consultada/os durante el desarrollo de este.

Referencias

- Bontridder, N., y Poulet, Y. (2021). El papel de la inteligencia artificial en la desinformación. *Datos y políticas*, 3(32), 1-21. <https://doi.org/10.1017/dap.2021.20>.
- Croitoru, F. A., Hondru, V., Ionescu, R. T. y Shah, M. (2023). Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9), 10850-10869. <https://doi.org/10.1109/TPAMI.2023.3261988>.
- García-Marín, D. (2022). Modelos algorítmicos y fact-checking automatizado. Revisión sistemática de la literatura. *Documentación de las Ciencias de la Información*, 45(1), 7-16. <https://doi.org/10.5209/dcin.77472>.
- Martín-Neira, J. I., Trillo-Domínguez, M. y Olvera-Lobo, M. D. (2023). Ibero-American journalism in the face of scientific disinformation: Fact-checkers' initiatives on the social network Instagram. *Profesional de la información / information Professional*, 32(5), e320503. <https://doi.org/10.3145/epi.2023.sep.03>.
- Parikh, S. B., y Atrey, P. K. (2018). Media-Rich Fake News Detection: A Survey. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 436-441.
- Porlezza, C. (2023). 'Promoting responsible AI: European perspective on the governance of artificial intelligence in media and journalism', *Communications*, 48:3, 370-394, <https://doi.org/10.1515/commun-2022-0091>.
- Sánchez, H. M., Sánchez, M. y Alonso, M. (2024). 'Artificial Intelligence and disinformation literacy programmes used by European fact-checkers'. *Catalan Journal of Communication & Cultural Studies*, 16(2), 237-255. https://doi.org/10.1386/cjcs_00111_1.
- Sánchez, H. M., Sánchez, M. y Martos, J. (2024). The methodology used by fact-checkers. An in-depth analysis of commonly used strategies. *Journalism Practice*, 1-24, <https://doi.org/10.1080/17512786.2024.2340522>.
- Sánchez, M., Sánchez, H. M. y Martínez, S. (2022). Inteligencia artificial en verificadores hispanos de la red IFCN: proyectos innovadores y percepción de expertos y profesionales. *Estudios sobre el Mensaje Periodístico*, 28(4), pp. 867-879. <https://doi.org/10.5209/esmp.82735>.
- Sánchez González, M., Sánchez Gonzales, H. M. y Martos Moreno, J. (2022b). Emprendimiento en verificación periodística: fact-checkers hispanos. En: Sánchez Gonzales, H. M. (ed.): *La digitalización en el periodismo. Transformación, retos y oportunidades* (145-168). Gedisa/ SEP.
- Sánchez González, M., Sánchez Gonzales, H. M. y Martos Moreno, J. (2022a). Innovación editorial en redes sociales de los verificadores hispanos de la #CoronavirusFact-Check Alliance: contenidos y visión de sus responsables. *Revista Latina de Comunicación Social*, 80, pp.135-161. <https://www.doi.org/10.4185/RLCS-2022-1535>.
- Shepherd, D.A. y Majchrzak, A. (2022). Machines augmenting entrepreneurs: Opportunities (and threats) at the Nexus of artificial intelligence and entrepreneurship. *Journal of Business Venturing*, 37(4), 106227. <https://doi.org/10.1016/j.jbusvent.2022.106227>.

La audiencia juvenil ante la desinformación: percepciones y reacciones en torno a la información política

Amparo Huertas Bailén

Institut de la Comunicació (InCom) – Universitat Autònoma de Barcelona
Amparo.huertas@uab.cat

Luiz Peres-Neto

Institut de la Comunicació (InCom) – Universitat Autònoma de Barcelona
Luiz.peres@uab.cat

1 | Introducción

El acceso a información de calidad es fundamental para el sostenimiento de las democracias. Si la participación ciudadana es la esencia de todo sistema democrático, es evidente que el derecho a recibir información de calidad ha de estar garantizado. Ahora bien, en el contexto digital actual, la credibilidad de los medios de comunicación y la veracidad informativa están en crisis (Rodríguez Borges, 2024; Lorenz-Spreen *et al.*, 2022). Las audiencias acceden hoy en día a un espacio digital confuso, donde los medios conviven con los pseudomedios, donde lo falso se presenta como verídico con el fin expandir mentiras entre la opinión pública y donde los mensajes conspiranoicos y la seudociencia tienen más recorrido por su capacidad de viralización.

Rodríguez Pérez (2019) habla de camuflaje de contenidos disfrazados de noticias y plantea el problema como un choque entre la libertad de información y la libertad de expresión. Este autor considera que ambas libertades son reo del «libertinaje del ciberespacio»: la primera, con la expansión de la desinformación y, la segunda, de la polarización.

Cuando el foco se coloca sobre la actualidad política, la suma de desinformación y polarización resulta un problema social todavía más preocupante si cabe, pues este tipo de contenidos está estrechamente ligado con la calidad del debate público y democrático (Grossi, 2007). Y si lo que se observa es la juventud, lo que se advierte es que este es un problema que puede ir agravándose con el tiempo, pues el consumo informativo de esta franja de edad se concentra en las redes sociales (Ceballos-del-Cid *et al.*, 2025), que es el lugar donde más cómodamente circula la desinformación y la discusión polarizada (Bail *et al.*, 2018; Vosoughi *et al.*, 2018). Y es también en esta franja de edad donde se concentran los más altos niveles de abstención de voto.

Además, tampoco podemos olvidar que ese período vital acoge una gama de edades cada vez más amplia, lo que agrava el problema como hecho social. La ONU considera como personas jóvenes las comprendidas entre los 15 y 24 años, tal y como manifestó en 1985 con motivo del Año Internacional de la Juventud y fue aprobado por la Asamblea General en su resolución 36/28 de 1981. Los estudios, estadísticos o no, que ahondan en su consumo y recepción mediática suelen ajustarse a esos márgenes, tal y como puede verse, por ejemplo, en una sistematización de las investigaciones realizadas sobre la relación entre autoestima y uso de redes sociales entre jóvenes (Campodónico y Aucapiña, 2024), donde se detecta que las franjas de edad más empleadas son: de 12 a 17 años, de 15 a 18 años y de 15 a 22 años. Ahora bien, cada vez es más habitual encontrar estudios sobre jóvenes que llegan a abarcar hasta los 29 años —o, incluso, más—. Un ejemplo en el contexto español es el «Barómetro Juvenil 2023. Salud y Bienestar» de la Fundación Mutua Madrileña y Fad Juventud. O, a finales de 2024, el Observatorio de la Juventud (INJUVE) incluyó en su encuesta por primera vez a las personas de edades comprendidas entre los 30 y los 34 años (Instituto de la Juventud, 2024).

Nuestro proyecto se ha desarrollado en Cataluña, donde disponemos de un sondeo que da cuenta de la actitud de la juventud ante la política. Es más, este fue realizado coincidiendo con parte de nuestro trabajo de campo, concretamente entre el 2 de mayo y el 29 de julio de 2023. Se trata de un informe sobre la participación e implicación política de la juventud catalana que surge del análisis de la última edición de la Encuesta sobre generaciones y participación (EGP), impulsada por el Centre d'Estudis d'Opinió de la Generalitat de Catalunya (Galais, 2024). Los datos referentes a las personas con edades comprendidas entre los 16 y los 29 años, agrupadas en el estudio como jóvenes, indican, efectivamente, que

las redes sociales son la principal vía para acceder a la información política. Entre los 20 y los 24 años, el 28,4% de la muestra así lo indica y, entre los 25 y 29 años, el índice baja ligeramente hasta el 26,4%. En segundo lugar, pero a gran distancia, aparece la televisión, con el 11,1% y el 12,0% respectivamente para cada una de estas dos franjas de edad. La tendencia a un equilibrio en el consumo de medios (prensa, televisión, radio y redes) según edad no aparece hasta los 45 años y, a partir de los 65 años, la televisión acaba siendo claramente la principal fuente informativa. En lo que se refiere al nivel de abstención, es en el colectivo juvenil donde aparece una tendencia al alza más marcada: entre 2017 y 2023, este índice ha pasado del 24,4% al 36,5%, según este mismo informe.

Sobre el consumo informativo, que es el eje central de este trabajo, conviene subrayar que todos los trabajos disponibles apuntan a un claro dominio de las redes sociales. Es decir, este es un fenómeno global que, además, aparece vinculado a otro proceso paralelo, el progresivo abandono de los medios tradicionales (Drulă, 2024; Fletcher, 2024; Scheffauer *et al.*, 2021). Ahora bien, no todas las redes son identificadas como informativas. Del conjunto de plataformas disponibles y siguiendo los datos de Reuters Institute (2024), entre las personas de 25 a 35 años destaca el uso de Instagram y X (antes Twitter). En relación con la franja etaria previa, de los 18 a 24 años, Reuters Institute (2024) detecta una tendencia al alza notable de la penetración de Tiktok.

En este capítulo nos acercamos al consumo informativo juvenil desde la mirada de la propia audiencia. Tradicionalmente, el estudio del consumo informativo se ha centrado básicamente en el análisis del grado de credibilidad o fiabilidad en los medios. Actualmente, con el desarrollo de las redes sociales, emergen nuevos objetos de estudio. Muchos de ellos giran en torno al potencial adictivo del entorno digital y ahondan en los efectos de la aplicación de los algoritmos. Lo que abordamos aquí es un tema específico, cómo la juventud se enfrenta a la desinformación en el terreno de la información política, un contenido que podemos afirmar con total seguridad que ya es imposible de eludir. Con el foco puesto en este asunto, nos interesa conocer: (a) ¿de qué modo identifican los agentes que distribuyen la desinformación?; (b) ¿de qué modo describen la desinformación?; (c) ¿de qué modo relacionan la curación algorítmica (las recomendaciones personalizadas) con la recepción de desinformación? y (d) ¿qué estrategias siguen para verificar la información que les llega? En términos generales, nuestro propósito es intentar explicar —o, al menos, intentar entender— de qué manera es percibida la desinformación por parte del colectivo juvenil (de 19 a 33 años) a partir del estudio de sus propios relatos sobre su propio consumo.

Este texto es fruto del proyecto de investigación «Polarización social e interculturalidad: el seguimiento de la actualidad política por parte de la juventud migrante y autóctona desde una mirada interseccional

(POINTAP)» (PID2021-125032OB-I00), financiado por el Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación /10.13039/501100011033/ y por FEDER «Una manera de hacer Europa».

2 | El consumo de información política desde la mirada de las audiencias

Una lectura de los estudios ya disponibles sobre la expansión de la desinformación y su consumo permite identificar tres ejes temáticos como los más abordados: (a) los patrones de propagación de la desinformación en entornos digitales derivados de la aplicación de estrategias algorítmicas; (b) la efectividad de los sistemas de verificación disponibles y (c) la recepción de contenidos generados por inteligencia artificial con un alto nivel de verosimilitud y, muy especialmente, de los videos conocidos como *deepfakes* que muestran imágenes falsas de personas (López López *et al.*, 2023).

Dado que nuestro trabajo se centra en el seguimiento de la actualidad política, también hemos explorado los estudios específicos en el área de esta especialización comunicativa. En este caso, los proyectos focalizan su interés en la polarización, es decir, en la libertad de expresión (Peres Neto y Huertas Bailén, 2024). La mayor parte se han realizado durante campañas electorales en Estados Unidos y, en el caso de estudiarse la responsabilidad o participación de los partidos políticos, el comportamiento de la extrema derecha ha sido el más analizado. Es decir, apenas han sido estudiados los momentos no coincidentes con elecciones y las prácticas político-comunicativas que se desarrollan en Estados Unidos tienden a consolidarse como una referencia que ayuda a prever lo que ocurrirá en otras zonas geográficas y, además, en muchas ocasiones, sin reflexión crítica. El comportamiento de los bots, es decir, el contenido generado por falsos perfiles, es una de las cuestiones que más preocupan. Esto no resulta extraño, pues hay claras pruebas de que estas estrategias han sido determinantes en elecciones de gran peso, como fue el caso del Brexit, el referéndum de 2016 que desencadenó la salida del Reino Unido de Europa (Kubin & Sikorski, 2021).

Nuestra propuesta parte de la suma de este conjunto de preocupaciones, pero se propone analizarlas desde la mirada de las audiencias para saber, por decirlo en términos generales, hasta qué punto inquietan también a la ciudadanía. El entramado digital ensalza el consumo hiperpersonalizado como gran avance. El sistema alardea de ser capaz de hacer llegar a cada usuario los contenidos que más le interesan a este, pero apenas existen trabajos que aborden esa experiencia desde la mirada de los receptores.

Pariser (2011) ya pronosticó que los usuarios de las redes sociales quedarían expuestos a una reducida gama de contenidos, contactos y opiniones. De hecho, conceptos derivados de esta perspectiva, como «efecto

burbuja» o «cámara de resonancia», han sido clave en el desarrollo de recientes teorías de comunicación (Peres Neto y Huertas Bailén, 2024). Además, desde la recepción, resulta difícil separar «libertad de información» y «libertad de expresión» cuando sabemos del fuerte vínculo de los medios de comunicación con la política, un fenómeno histórico ampliamente debatido. Hallin & Mancini (2004) hablan abiertamente de «pluralismo polarizado» para referirse a la militancia política explícita de los medios de comunicación y sitúan a España como uno de los países más afectados. Estos son solo algunos ejemplos de teorías y paradigmas con largo recorrido en el ámbito de la comunicación, pero que claramente están más pensadas desde la mirada de los productores y las instituciones políticas que desde la perspectiva de la audiencia.

2.1. Los rastros digitales como herramienta de estudio

La convivencia de medios de comunicación con otras fuentes ajenas a la práctica periodística en un mismo entorno digital complica el estudio del consumo. El alto volumen de contenidos en circulación y la diversidad de procedencias hace que clasificaciones que ya fueron creadas atendiendo al nuevo contexto sociotecnológico, como las de Chadwick *et al.* (2018), hayan dejado de ser operativas. Estos autores distinguían entre circulación directa (de los medios a las audiencias) e indirecta (el contenido que comparten los usuarios en las redes sociales). Esta distinción entre medios y audiencias nos resulta hoy en día incluso ingenua, dados los múltiples agentes, con intereses diversos, que ahora intervienen en el ciberespacio, desde los pseudomedios hasta los partidos políticos pasando por los influencers.

Según Salazar Rebolledo (2024), la estructura de comunicación «no profesional» que prolifera en el ámbito digital ha anulado la influencia de los profesionales periodistas. Para esta autora, esta suplantación del mediador periodístico por agentes ajenos a la práctica profesional ha sido clave en la expansión de la desinformación. Ekströmy & Westlund (2019) hablan de dislocación informativa.

En este contexto, es lógico pensar que a la ciudadanía le resulte difícil identificar y recordar las fuentes de información utilizadas, lo que entorpece el estudio de su consumo cuando se opta por trabajos empíricos basados en la consulta del recuerdo a muestras poblacionales. Y, en el caso de la población juvenil, la cuestión se perfila como especialmente delicada, pues sus prácticas comunicativas se han formado en un entorno completamente digital, donde, por ejemplo, la búsqueda de medios ha dejado de tener sentido, pues el sistema se fundamenta en las búsquedas de temas. Además, este colectivo se ha habituado a tener un acceso a la información inmediato, directo y gratuito. Es decir, la recepción juvenil está muy alejada del modelo tradicional.

En este marco, las tendencias actuales del estudio del consumo informativo se concentran en el análisis de los rastros digitales. La facilidad para obtener datos en el contexto digital ha contribuido en ello. Los equipos de investigación tenemos a nuestro alcance *software* que nos permite descargar fácilmente un ingente volumen de información y mensajes compartidos en redes. Y, por otra parte, la consolidación del vídeo como formato estrella está haciendo que los estudios de medición cada vez se enfoquen más en el análisis de la interactividad que provoca cada vídeo. De hecho, y como consecuencia del funcionamiento de Tiktok —una red que no se basa en el contacto con personas próximas o conocidas—, ya comienza a plantearse que el número total de seguidores de una cuenta dejará de ser pertinente para medir su éxito una vez que puede ser más efectivo el análisis de la popularidad de cada vídeo, medida a partir de la interacción que este genera (Gerbaudo, 2024; Newman, 2022). Desde el punto de vista teórico, esto supone asumir que los algoritmos se han convertido en los verdaderos curadores del consumo (Beaumont, 2022; Lopezoza *et al.* 2023; Boté-Vericad, Adilović y Afonso-Mendonça, 2024).

En el área del consumo de noticias, dos temas emergen con fuerza como objeto de estudio. En primer lugar, la recepción involuntaria de noticias o consumo incidental (Farias-Batlle *et al.*, 2024; Scheffauer *et al.*, 2021; Chen *et al.*, 2021). En segundo lugar, el rechazo a las noticias, un problema que afecta sobre todo a las generaciones más jóvenes (Díaz-Noci y Pérez-Altable, 2023; Skovsgaard & Andersen, 2020; Goyanes *et al.* 2023; Mellado y Cruz, 2024).

El hecho de que las redes sociales sean un campo abonado para la desinformación y que ese sea el lugar de mayor consumo juvenil hace que domine una mirada pesimista en la investigación disponible, muy especialmente en aquellos estudios que no se centran en el potencial de la educación mediática. No obstante, también aparece alguna mirada optimista, como la de Kruikeimer & Shehata (2017). Estos autores señalan que la retroalimentación entre redes y otros medios, como la televisión, puede acabar propiciando la participación y el interés hacia la política. O, un estudio realizado en Estados Unidos sobre la relación entre afinidad político-partidaria y aceptación de vacunas contra la COVID-19 identificó que, en sentido contrario a lo que es habitual leer, las redes sociales polarizaron menos que los medios tradicionales (Jones-Jang & Chung, 2022).

2.2. La defensa del estudio de los relatos de las audiencias

Este segundo apartado lo queremos dedicar a defender el estudio de los públicos de forma directa, y no únicamente a través de los rastros digitales. Además, gran parte de los trabajos disponibles han sido definidos desde la perspectiva del productor y, por tanto, creemos pertinente

defender la necesidad de más propuestas que no solo se aproximen a la audiencia, sino que también asuman su mirada. Duvekot *et al.* (2024), en una revisión sistematizada de los textos publicados entre 2006 y 2022 sobre el consumo de información por parte de personas de 10 a 36 años, tan solo registró 94 artículos basados en trabajos empíricos con muestras poblacionales.

Son ya varios los autores que manifiestan el problema que puede derivarse de la sobrevaloración de lo digital, como es el caso de Axel Burns (2019), quien considera que conceptos como «cámaras de eco» o «filtros burbuja» son demasiado simplistas, pues solo responden a la lógica tecnológica cuando el proceso es mucho más complejo. O Livingstone (2019), quien ya señaló que los estudios basados en rastros digitales ofrecen datos parciales y descontextualizados, una preocupación que compartimos.

Nos alineamos así con el planteamiento de Swart *et al.* (2022), quienes defienden la consideración de las audiencias como un agente activo más en el engranaje comunicativo. En concreto, estos autores plantean la pertinencia de desarrollar enfoques no centrados exclusivamente en los medios de comunicación y cubrir todas las vías desde donde se distribuye contenido informativo, la inclusión de todo tipo de audiencias y no solo de los perfiles con mayor potencial comercial y, por último, ahondar en la experiencia de los públicos.

Pero, además, nuestra propuesta defiende de manera muy especial el uso de metodologías cualitativas. En este sentido, nos interesan especialmente las ideas de Hendrickx (2024), quien apunta que los trabajos empíricos cuantitativos no deberían pretender entender qué motiva a los públicos a seguir un determinado comportamiento, pues el análisis numérico no aporta detalles que permitan deducir esa información, mientras que las propuestas cualitativas, por el contrario, sí pueden ayudar a elaborar ese tipo de interpretaciones. Compartimos la mirada de Hendrickx (2024), también por su optimismo. Este autor percibe que se está dando un auge del uso de metodologías cualitativas y, además, su planteamiento no tiene como finalidad entrar en conflicto entre diferentes «culturas de investigación» sino, todo lo contrario, defender su complementariedad.

De momento, son escasas las aportaciones académicas centradas exclusivamente en la experiencia y visión del consumidor en entornos digitales que atiendan a sus discursos y argumentos. Pero, por su originalidad, queremos destacar una de ellas. Lomborg & Kapsch (2020) se propusieron indagar sobre la experiencia algorítmica de las audiencias y, más concretamente, sobre su nivel de conocimiento del funcionamiento de los algoritmos, los sentimientos que estos provocan y la evaluación de los mismos. A partir de 16 entrevistas en profundidad, Lomborg & Kapsch (2020) acabaron negando que los algoritmos sepan más de los

usuarios que ellos mismos y detectaron aspectos muy interesantes como, por ejemplo, un gran rechazo por parte de las personas de los colectivos LGTB ante el contenido que les era recomendado, pues en la mayoría de las ocasiones no tenían en cuenta ni su género ni su orientación sexual.

3 | Metodología

Nuestro trabajo de campo, realizado en Cataluña entre 2022 y 2024, se ha basado en la realización de entrevistas en profundidad a una muestra total formada por 50 personas autóctonas y migrantes (con edades comprendidas entre los 19 y los 33 años). Pero, para abordar este capítulo, nos hemos centrado exclusivamente en la submuestra de personas nacidas en España sin experiencia migratoria (n: 23).

A continuación, detallamos las características sociodemográficas de la muestra consultada:

- Según el lugar de nacimiento, la mayoría (18) son personas nacidas en Cataluña y el resto responde a procesos de migración interna: 2 proceden de Andalucía; 1, de Baleares; 1, de Valencia y 1, de Galicia.
- Según el sexo, la muestra está constituida por 8 hombres y 15 mujeres.
- Según el género/orientación sexual, 13 son heterosexuales y 10 responden a identidades no normativas (3 hombres homosexuales, 6 mujeres homosexuales y 1 mujer trans).
- Según la edad, 13 pertenecen a la franja que comprende de los 20 a los 25 años y 10, a la que se sitúa entre los 26 y los 29 años.
- Según la clase social, la mayoría (16) pertenecen a la clase media, mientras que 5 se sitúan en la clase alta y 2, en la clase baja.
- Según la situación laboral, 11 trabajan, 8 estudian y 4 están en situación de desempleo.
- Sobre sus núcleos de convivencia, las situaciones son muy variadas: 10 viven con la familia; 6, con la pareja sentimental y el resto (7) presentan situaciones diversas (piso compartido con estudiantes, piso compartido con personas con quienes no se tiene relación, alquiler de una habitación, ...).
- Según la ideología política, 3 se posicionan en la extrema izquierda; 13, en la izquierda; 4, en el centro; 1, en la derecha y 1, en la extrema derecha. Además, una de las personas entrevistadas no se define de forma concreta (su posicionamiento ideológico en el eje derecha-izquierda depende del tema que se trate).

Es importante resaltar el contexto en el que se ha realizado este estudio, tanto el periodo temporal como la comunidad autónoma. Además

de las elecciones generales ocurridas en el territorio español entre 2022 y 2024, se ha de tener presente la peculiaridad de Cataluña, una comunidad donde la defensa de la independencia del Estado español ha marcado fuertemente su agenda política.

El guion de las entrevistas se divide en 4 bloques. Mientras que el primero se centra en la historia personal con especial énfasis en la participación política y el grado de implicación en movimientos sociales en el seno familiar y en los círculos amistosos previos, los otros tres apartados abordan de forma directa los aspectos que definen el objeto de estudio del proyecto del que surge este texto centrándose ya en el presente. Estos son: socialización política; actitud, valores y posicionamiento ideológico y, por último, seguimiento de la actualidad informativa-política. Las entrevistas duraron entre 30 y 80 minutos, dependiendo del volumen de información a aportar por cada persona participante.

Ahora bien, para el desarrollo de este capítulo hemos analizado únicamente la información disponible sobre su percepción y reacciones ante la desinformación. En lo que se refiere a la percepción, nos interesa saber dónde la identifican, cómo la describen y si la asociación con la experiencia algorítmica. En relación con las reacciones, nos proponemos analizar si acostumbran a verificar los contenidos que sospechan pueden ser falsos o sesgados.

4 | **Discusión y resultados**

4.1. **Introducción: Panorámica general del consumo de información política**

El total de la muestra expresó estar interesada en seguir la información de actualidad, 14 presentaron un seguimiento constante y 9, con un carácter menos regular. En relación a los *influencers*, 13 indicaron seguir a personas que abordan temas políticos, con gran peso del contenido feminista (especialmente con toque humorístico y entre la muestra de mujeres lesbianas).

Sobre las principales fuentes de información, 19 expresaron emplear internet y, muy especialmente, las redes. Del resto de la muestra, cabe destacar que solo un perfil con alto interés informativo recurre exclusivamente a los medios tradicionales, siguiendo directamente la prensa y la radio informativa (dentro y fuera del entorno digital), y los otros tres restantes no plantean ningún medio como destacado en consonancia con el hecho de que son personas que no tienen un hábito regular de consumo informativo.

No obstante, este dominio de las redes sociales como fuente de información no puede interpretarse como si se tratara de un uso exclusivo o incompatible con otros medios. De las 19 personas que mencionaron

internet/redes, solo 4 (3 con alto interés informativo) concentran su consumo aquí. Por tanto, lo que domina es la hibridación de redes con medios tradicionales. Por ejemplo, 6 personas combinan redes con televisión y prensa (5 son mujeres); 5, con prensa específica (4 son hombres) y 4, con televisión. De hecho, 13 personas afirmaron seguir la prensa y 12, la televisión. En cambio, la penetración de la radio es apenas perceptible, tan solo fue mencionada en 2 entrevistas.

Conviene subrayar también que muchas de las personas entrevistadas apuntaron sentir que las redes les ocupan demasiado tiempo en su día a día. La gran parte de la muestra no está dispuesta a mantener muchos perfiles activos en diferentes redes, ni tampoco a experimentar con nuevas plataformas que puedan aparecer. De hecho, es habitual plantear el interés por reducir el uso a solo dos plataformas.

Para cerrar este breve resumen del interés por la actualidad política y de la dieta mediática de la muestra consultada, creemos preciso incluir aquí una aclaración. Esta tiene que ver con un comportamiento que descubrimos durante el propio transcurso de las entrevistas y que nos resultó llamativo. Tras informar del objetivo de nuestro estudio, los miembros de la muestra solían necesitar un momento de reflexión para centrarse en la propuesta. De alguna manera necesitaban un instante para poder diferenciar entre su consumo de información y de entretenimiento y, a partir de ahí, poder concentrarse solo en lo primero. Consideramos que este no es un tema menor, pues en cierto modo revela que la población observa ambos consumos claramente diferenciados y, de hecho, para solventar esta cuestión, la tendencia mayoritaria fue asignar a cada red social una función concreta y diferenciadora. Así, mientras que Instagram destacó como una red habitual como espacio digital de búsqueda de información, Twitter generó múltiples críticas al asociarse con una excesiva polarización y Tiktok apenas fue mencionada, pues mayoritariamente fue vinculada con el consumo de entretenimiento y/o con franjas de edad por debajo de los 19 años.

4.2. Medios de comunicación y redes sociales: percepciones diferentes

La muestra consultada percibe más desinformación en los medios de comunicación tradicionales que en las redes sociales. Ahora bien, para entender esta idea, hay que explicar que lo que detectamos exactamente es que las prácticas periodísticas de los medios tradicionales son identificadas por la muestra como causa directa de desinformación. Los argumentos giran o bien en torno a una excesiva presencia de la información institucional o bien en torno a la idea de una información sesgada por

estar excesivamente posicionada políticamente. Los discursos tejidos en este sentido aluden al silenciamiento o ausencia de los temas que quedan fuera de la política institucional y a información incompleta por no recoger todos los puntos de vista.

La agenda mediática está protagonizada por la agenda institucional y es que paso (...) no quiero estar leyendo 25 noticias y que 20 sean de partidos políticos. A mí me gustaría escuchar sobre lo que pasa en x lugar, y no hace falta irse muy lejos... qué está pasando en este lugar, en esta fábrica... cuáles son los problemas laborales de estas personas (...) Lo que pasa en Francia... estamos muy pendientes de lo que pasa en Francia a nivel institucional, lo entiendo porque es lo que marca, pero creo que es una manera de eludir la responsabilidad cuando tienen la responsabilidad de lo que nos hacen creer... Me interesan más las cosas que tienen menos altavoz (E16: Hombre heterosexual, 22 años, traducido del catalán).

Entonces cuando miro las noticias (en televisión) pienso... es que todo es según el punto de vista desde que te lo mires, ¿no? (E20: Mujer heterosexual, 29 años, traducido del catalán).

En cambio, el espacio digital se vincula con el pensamiento crítico. La muestra consultada siente que el hecho de poder realizar búsquedas personalizadas o poder decidir qué perfiles seguir y cuáles descartar les permite elegir más libremente qué contenidos consultar y eso lo asocian con el acceso a una información más abierta.

No obstante, también hay que tener en cuenta que 8 componentes de la muestra afirmaron seguir el canal informativo autonómico catalán 3/24 a través de Instagram, 7 son mujeres y 1 es un hombre. Es decir, por un lado, las redes también suponen un acceso a los medios tradicionales y, por otro, todavía es posible hablar de fidelización en el consumo.

Atendiendo a la muestra consultada, la audiencia valora muy positivamente el volumen y diversidad de fuentes que internet le ofrece. Incluso, en este contexto de búsqueda y recepción, cuando aparece contenido procedente de los medios de información tradicionales este no parece provocar rechazo.

me salen noticias en Instagram (se refiere a noticias de diarios como *Ara* o *El Periódico*) y entonces una noticia que me genera curiosidad o me interesa, entro al diario (E06: Mujer lesbiana, 27 años, traducido del catalán).

Evidentemente, esos contenidos que les llegan son resultado de una estrategia algorítmica, pero no siempre este proceso es identificado. Para ilustrar esta idea, resulta especialmente llamativo el siguiente caso:

si tú tienes Instagram, Instagram va a adaptarse a tu forma de pensar, YouTube lo mismo. Microsoft Edge hasta ahora, no. Microsoft me pone todo, de todas las opiniones. Y muchas veces me leo artículos que a mí me parecen una barbaridad por lo que me están diciendo (por reflejar una ideología política distinta a la propia) y, en ese sentido, me gusta porque siento que no se está adaptando el algoritmo a mí... coge artículos *random*. En plan, ahora te coge un artículo de «diario.es», ahora uno de «La Vanguardia», ahora uno de «Expansión», ahora uno de «El Confidencial», del «Huft Post»... Miro todas las (noticias) que hay, te salen 10 principales, y miro todas las que hay. (...) Pero todo lo que es política, lo miro todo (E10: Hombre heterosexual, 23 años).

4.3. Actuaciones frente a la desinformación

Entre la muestra consultada, destacan tres actuaciones que tienen como objetivo comprobar la veracidad de la información: (a) consultar las fuentes originales; (b) ampliar la información accediendo a fuentes alternativas o diferentes a la que ha sido vehículo de lo que se percibe como desinformación y (c) verificar en medios lo que han recibido de fuentes no periodísticas.

El consultar las fuentes originales se perfila como una práctica iniciada a raíz de la COVID-19. Algunos miembros de la muestra remiten que durante el confinamiento motivado por aquella pandemia se acostumbraron a revisar las cifras de fallecimientos en las webs oficiales del Ministerio de Sanidad. No obstante, la idea de recurrir a fuentes originales puede remitir a recursos muy diversos, no solo a documentos completos oficiales.

lo que utilizo principalmente es Twitter, donde me van apareciendo noticias de todo tipo de medios de comunicación, de todos los campos. Entonces, yo no suelo entrar en las noticias, porque ahora hay unos periódicos por los que tienes que pagar... y no voy a pagar... pero intento siempre tener la versión de diferentes titulares (...) y decir esto no tiene ni el más mínimo sentido. Pero normalmente por lo que me informo son por declaraciones... es decir, los videos de esta rueda de prensa... esta declaración institucional... este tipo de cosas... por ejemplo, durante el confinamiento tampoco me interesaba mucho ver el tema de las noticias porque ver durante una hora... los muertos de no sé dónde... los de tal... el hospital no sé cuánto... no me interesaba... entonces yo lo que hacía era... entrar en el Ministerio de Sanidad, descargarme el PDF y decir, vale, cómo ha cambiado la gráfica respecto a ayer... soy una persona que prefiere el dato. Es decir, tengo el dato y con los datos no te pueden mentir (E14: Hombre homosexual, 26 años).

Por su clara cercanía con la actualidad política, otra tendencia que queremos destacar aquí es la consulta directa de los programas de los partidos políticos durante las campañas electorales.

miramos las webs (en familia) porque (los programas) son tan similares y quizá (en la decisión de voto) se depende demasiado de la persona que los representa (si no se consultan los programas). Sí, entonces miro las webs. Las últimas dos veces he votado por tema webs. Léí las ideas, porque si nos hemos de fiar de los 2 segundos que salen en las noticias, es que no, noooo (E20: Mujer heterosexual, 29 años, traducido del catalán).

Yo soy bastante apolítica en verdad y, entonces, no me posiciono fuerte en ninguna ideología política, ni en ningún... digamos... en ninguna creencia política al 100%. Cuando hay que votar y eso, no voto por partido, sino por programa. Entonces, hay programas que algunas partes me parecen mejor que otras... pero, al final, voto como por lo menos malo, según yo pienso. No por el partido (E27: Mujer heterosexual, 24 años).

Sobre la costumbre de comprobar si una información es real consultando si aparece en los medios, esta entrevista resulta un buen ejemplo:

Hago una cata de vídeos, y cuando ya tengo el titular o la noticia de la que quiero saber más, lo busco (...) Puedes coger una noticia y después buscar un diario o buscar la fuente (...) Los *trending topics* también los miro... como hace dos días... el director de cine Carlos Vermut (acusado de acoso sexual) salió en El País, que ha abusado de tres mujeres, ha sido acusado por 15 personas que forman parte del mundo de la farándula (E29: Mujer lesbiana, 27 años, traducido del catalán).

Conclusiones

A pesar de las limitaciones de los estudios cualitativos, pues por su propia naturaleza solo permiten trabajar con pequeñas muestras poblacionales, entendemos de gran utilidad poder conocer las experiencias individuales, la evolución de los hábitos relacionados con la búsqueda de información y los motivos que llevan a los individuos a tomar decisiones que implican cambios. Como ha podido comprobarse, nuestro estudio reproduce los datos de consumo que reflejan los grandes sondeos representativos estadísticamente —nos referimos sobre todo al dominio de las redes sociales—, pero también ayuda a entender cuestiones como el vínculo que la ciudadanía establece entre los distintos medios o la funcionalidad que asigna a cada ámbito o plataforma. Consideramos que esta información es muy útil para interpretar los datos numéricos de los grandes sondeos.

El trabajo apunta diversas cuestiones de interés. De todas ellas, queremos destacar dos. En primer lugar, el hecho de que continúa siendo posible hablar de fidelidad a los medios, como resulta con el canal informativo autonómico 3/24, cuya estrategia en Instagram está despertando el interés entre la juventud. Y, en segundo lugar, la importancia de la transparencia informativa institucional. Es interesante pensar que la consulta directa de los programas electorales de cada partido o la búsqueda de información en páginas web de la administración puede estar siendo una vía de salida en respuesta a la desaparición del papel de la prensa como mediadora, pero también puede verse en sentido positivo, ya que refleja una ciudadanía implicada.

Al mismo tiempo, este trabajo nos está haciendo pensar sobre qué significa «estar informado» hoy en día, ante los retos que suponen los fenómenos de la desinformación y de la polarización. Por alejarse del objetivo del capítulo, no se ha abordado el hecho de que mayoritariamente la juventud se interesa por temas muy concretos, por aspectos que suelen en muchos casos acabar siendo el eje central de sus proyectos de vida, como el cuidado del medio ambiente, la defensa del derecho a la vivienda o la lucha feminista. Por tanto, aunque podríamos decir que su mirada sobre el entorno es muy cerrada, se ha de tener en cuenta que muchas veces va vinculada al compromiso social.

Pero trabajos como este nos acaban, en realidad, generando más preguntas que respuestas, por ejemplo ¿a partir de qué nivel de conocimiento una persona siente estar informada sobre un tema? Esperemos que la suma de nuevos proyectos en línea con lo que aquí hemos propuesto nos permita seguir avanzando. Este ha de ser un esfuerzo colectivo de los equipos de investigación.

Referencias

- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F. y Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221. <https://doi.org/10.1073/pnas.1804840115>.
- Beaumont, R. (2022). IA é o novo DJ: Inteligência Artificial como curadora cultural. *Revista ALCEU*, 22(46), 48-55. <https://doi.org/10.46391/ALCEU.v22.ed46.2022.281>.
- Boté-Vericad J. J., Adilović, E. y Afonso-Mendonça, H. S. (2024). Agregadores de noticias: una aproximación al estado del arte sobre las consideraciones éticas en los algoritmos de recomendación. *Documentación de las Ciencias de la Información*, 47, 5-13. <https://doi.org/10.5209/dcin.90547>.
- Bruns, A. (2019). *Are filter bubbles real?* Polity Press.
- Campodónico, N. y Aucapiña, I. E. (2024) Revisión sistemática sobre la influencia de las redes sociales en la autoestima de los adolescentes. *Psicología UNEMI*, 8(15), 73-87. <https://doi.org/10.29076/issn.2602-8379vol8iss15.2024pp73-87p>.

- Ceballos-del-Cid, Y., Gómez-Calderón, B. y Córdoba-Cabús, A. (2025). Redes sociales y hábitos de consumo informativo de los jóvenes españoles: Un análisis diacrónico (2021-2023). *Revista Mediterránea de Comunicación*, 16(1), e28010. <https://doi.org/10.14198/MEDCOM.28010>.
- Chadwick, A., Vaccari, C. y O'Loughlin, B. (2018). Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing. *New Media & Society*, 20(11), 4255-4274. <https://doi.org/10.1177/1461444818769689>.
- Chen, H.-T., Kim, Y. y Chan, M. (2021). Just a glance, or more? Pathways from counter-attitudinal incidental exposure to attitude (de)polarization through response behaviors and cognitive elaboration. *Journal of Communication*, jqab046. <https://doi.org/10.1093/joc/jqab046>.
- Díaz-Noci, J. y Pérez-Altable, L. (2023). Por qué evitamos las noticias. Algunas evidencias desde la desigualdad de acceso a la información digital en España. *Anuario ThinkEPI*, 17, e17a34.h.
- Drulă, G. (2024). *Algorithms for online news production and audience*. CIAC Editions. <https://doi.org/10.34623/1v4s-cs86>.
- Duvekot, S., Valgas, C. M., de Haan, Y. y de Jong, W. (2024). How youth define, consume, and evaluate news: Reviewing two decades of research. *New Media & Society*, 0(0). <https://doi.org/10.1177/14614448241262809>.
- Ekström, M. y Westlund, O. (2019). The dislocation of news journalism: A conceptual framework for the study of epistemologies of digital journalism. *Media and Communication*, 7(1), 259-270. <https://doi.org/10.17645/mac.v7i1.1763>.
- Farias-Batlle, P., Córdoba-Cabús, A. y Gómez-Calderón, B. (2024). Jóvenes y redes sociales: Hábitos de consumo informativo y credibilidad de las noticias. *Comunicar*, 32(78). <https://doi.org/10.58262/V32I78.13>.
- Fletcher, R. (2024). Actitudes hacia los algoritmos e impacto en las noticias en Reuters Institute for the Study of Journalism (2023). *Digital News Report* (32-35). Reuters Institute. bit.ly/3EALADo.
- Galais, C. (Coord.) (2024). Participació i implicació política de la joventut catalana. Una anàlisi de l'Enquesta sobre generacions i participació de 2023. Departament de Drets Socials i Inclusió. bit.ly/41h5RHg.
- Gerbaudo, P. (2024). TikTok and the algorithmic transformation of social media publics: From social networks to social interest clusters. *New Media & Society*, <https://doi.org/10.1177/14614448241304106>.
- Goyanes, M., Ardèvol-Abreu, A. y Gil de Zúñiga, H. (2023). Antecedents of news avoidance: Competing effects of political interest, news overload, trust in news media, and 'News Finds Me' perception. *Digital Journalism*, 11(1). <https://doi.org/10.1080/21670811.2021.1990097>.
- Grossi, G. (2007). *La opinión pública: Teoría del campo demoscópico*. Madrid: CIS.
- Hallin, D. y Mancini, P. (2004). *Comparing Media Systems Three Models of Media and Politics*. Cambridge University Press.
- Hendrickx, J. (2024). 'Normal news is boring': How young adults encounter and experience news on Instagram and TikTok. *New Media & Society*, 0(0). <https://doi.org/10.1177/14614448241255955>.

- Instituto de la Juventud (2024). La juventud en cifras: población. *Injuve*. <https://acortar.link/wOgP9q>.
- Jones-Jang, S. M. y Chung, M. (2024). Can we blame social media for polarization? Counter-evidence against filter bubble claims during the COVID-19 pandemic. *New Media & Society*, 26(6), 3370-3389. <https://doi.org/10.1177/14614448221099591>.
- Kubin, E. y von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, 45(3), 188-206. <https://doi.org/10.1080/23808985.2021.1976070>.
- Kruikemeier, S. y Shehata, A. (2017) News Media Use and Political Engagement Among Adolescents: An Analysis of Virtuous Circles Using Panel Data. *Political Communication*, 34(2), 221-242. <https://doi.org/10.1080/10584609.2016.1174760>.
- Livingstone, S. (2019). Audiences in an Age of Datafication: Critical Questions for Media Research. *Television & New Media*, 20(2), 170-183. <https://doi.org/10.1177/1527476418811118>.
- Lomborg, S. y Kapsch, P. H. (2020). Decoding Algorithms. *Media, Culture & Society*, 42(5), 745-761. <https://doi.org/10.1177/0163443719855301>.
- López López, P. C., Mila Maldonado, A. y Ribeiro, V. (2023). La desinformación en las democracias de América Latina y de la península ibérica: De las redes sociales a la inteligencia artificial (2015-2022). *Uru: Revista De Comunicación Y Cultura*, (8), 69-89. <https://doi.org/10.32719/26312514.2023.8.5>.
- Lopezosa, C., Guallar, J., Codina, L. y Pérez-Montoro, M. (2023). Curación de contenido y periodismo: revisión sistematizada exploratoria y visión experta. *Revista Mediterránea De Comunicación*, 14(1), 205-223. <https://doi.org/10.14198/MEDCOM.22810>.
- Lorenz-Spreen, P. Oswald, L., Lewandowsky, S. y Hertwig, R. (2022). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, 7(1), 74-101. <https://doi.org/10.1038/s41562-022-01460-1>.
- Mellado, C. y Cruz, A. (2024). Navegando la sobrecarga informativa: Un estudio de distintos tipos de evitación de noticias entre las audiencias en Chile. *Cuadernos. Info*, (59), 48-70. <https://doi.org/10.7764/cdi.59.76243>.
- Newman, N. (2022). *How publishers are learning to create and distribute news on TikTok*. Reuters Institute for the Study of Journalism. bit.ly/4b1Nklm.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.
- Peres Neto, L. y Huertas Bailén, A. (2024). Pensar la relación entre comunicación y polarización política desde las epistemologías algorítmicas. *Lecciones del Portal de la Comunicación (InCom-UAB)*, Universitat Autònoma de Barcelona. bit.ly/4aZlsgw.
- Reuters Institute (2024). *Digital News Report*. University of Oxford. bit.ly/3ER6s9u.
- Rodríguez Borges, R. F. (2024). Periodismo en la era de la posverdad la UE al rescate de la información veraz, independiente y plural. *Intracom Journal*, 54-64. <https://doi.org/10.61283/mxprkn5655>.
- Rodríguez Pérez, C. (2019). No diga fake news, dí desinformación: una revisión sobre el fenómeno de las noticias falsas y sus implicaciones. *Comunicación*, (40), 65-74. <https://doi.org/10.18566/comunica.n40.a05>.

- Salazar Rebolledo, G. (2024). Votar en tiempos de desinformación: nuevos patrones informativos y comportamiento electoral. *Estudios sociológicos*, 42, <https://doi.org/10.24201/es.2024v42.e2538>.
- Scheffauer, R., Goyanes, M. y Gil de Zúñiga, H. (2021). Beyond social media news use algorithms: How political discussion and network heterogeneity clarify incidental news exposure. *Online Information Review*, 45(3), 633-650. <https://doi.org/10.1108/OIR-04-2020-0133>.
- Skovsgaard, M. y Andersen, K. (2020). Conceptualizing news avoidance: Towards a shared understanding of different causes and potential solutions. *Journalism Studies*, 21(4), 459-476. <https://doi.org/10.1080/1461670X.2019.1686410>.
- Swart, J., Groot Kormelink, T., Costera Meijer, I. y Broersma, M. (2022). Advancing a radical audience turn in journalism. Fundamental dilemmas for journalism studies. *Digital Journalism*, 10(1), 8-22. <https://doi.org/10.1080/21670811.2021.2024764>.
- Vosoughi, S., Roy, D. y Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>.

Inteligencia artificial, infoxicación e identificación de la desinformación por parte de jóvenes andaluces

Aránzazu Román-San-Miguel

Universidad de Sevilla

arantxa@us.es

Marta Sánchez-Hunt

Universidad de West London por Centro Universitario EUSA Sevilla

marta.hunt@eusa.es

1 | Introducción

El avance de la inteligencia artificial (IA) ha transformado profundamente los modos en que los jóvenes acceden y consumen información. Aplicaciones como motores de búsqueda inteligentes, asistentes virtuales y chatbots han facilitado el acceso rápido y personalizado a datos, convirtiéndose en herramientas comunes entre los universitarios. Sin embargo, esta dependencia plantea inquietudes sobre el impacto de la IA en la calidad de la información consumida, especialmente en cuanto a la capacidad de los jóvenes para distinguir entre contenido veraz y desinformación.

La desinformación ha proliferado con la creciente popularidad de las redes sociales como principal fuente de noticias para los jóvenes. Según el informe Media Navigator de Kantar, empresa líder mundial en consultoría e información basada en evidencia, un 71% de los jóvenes

entre edades de 16 y 24 años utiliza redes sociales como su fuente principal de información, mientras que solo el 35% recurre a medios tradicionales como periódicos o sitios web (Kantar, 2021). Además, solo uno de cada tres jóvenes puede identificar correctamente la fuente de la información que consume, lo que revela una preocupante falta de alfabetización mediática en este grupo etario. Este fenómeno, agravado por el uso de algoritmos de IA que priorizan contenido basado en el historial de búsqueda y las preferencias personales, contribuye a la creación de «burbujas informativas», donde los usuarios pueden estar expuestos principalmente a contenidos que confirman sus creencias, aumentando el riesgo de desinformación.

Dado que las plataformas de IA son cada vez más influyentes en la forma en que los jóvenes acceden a la información, resulta fundamental comprender cómo interactúan con ellas y cómo estas herramientas impactan su percepción de la realidad. Este estudio se centra en analizar el uso de las aplicaciones de IA entre los jóvenes universitarios andaluces de 18 a 20 años y en evaluar el grado de desinformación al que están expuestos.

A través de un enfoque de encuestas se analizan no solo los patrones de uso, sino también cómo los algoritmos de IA pueden contribuir a la creación de sesgos informativos y a la propagación de noticias falsas o descontextualizadas.

Este artículo tiene como objetivo contribuir al diseño de estrategias educativas y tecnológicas que promuevan un uso más crítico y consciente de las herramientas basadas en IA, minimizando los riesgos asociados a la desinformación en la era digital.

Este estudio se centra en investigar el uso de la inteligencia artificial (IA) entre jóvenes de 18 a 20 años y examinar el grado de desinformación que puede surgir a partir del uso de estas tecnologías para la obtención de información. En la actualidad, las aplicaciones basadas en IA, como los asistentes virtuales, chatbots y motores de búsqueda inteligentes, se han convertido en fuentes habituales para acceder a información. Sin embargo, su impacto en la comprensión y percepción de la realidad de los usuarios jóvenes sigue siendo un área poco explorada.

La investigación se desarrolla a través de un enfoque mixto, utilizando encuestas y entrevistas para recopilar datos sobre los patrones de uso de la IA en este grupo de edad. Además, se examina la influencia de estas aplicaciones en la generación de posibles sesgos informativos y en la proliferación de noticias falsas o descontextualizadas. Se busca no solo determinar la frecuencia y forma de uso de la IA por parte de los jóvenes, sino también identificar las principales fuentes de desinformación y los factores que contribuyen a su difusión.

Los hallazgos de este estudio permiten comprender cómo la interacción con la IA afecta el proceso de adquisición y validación de información

entre los jóvenes. Asimismo, los resultados contribuyen al diseño de estrategias educativas y tecnológicas que fomenten un uso más crítico y consciente de las herramientas de IA, minimizando así los riesgos asociados a la desinformación en la era digital.

2 | Inteligencia artificial y desinformación

2.1. De las redes sociales a la inteligencia artificial

El tránsito de las redes sociales a la IA en la obtención de información ha redefinido cómo los jóvenes interactúan con los medios. Las redes sociales, inicialmente vistas como plataformas de interacción social, se han convertido en fuentes predominantes de noticias. Aquellos que consideran a las redes sociales como un medio alternativo a los convencionales confían más en los contenidos y en las recomendaciones de los usuarios que participan en ellas (De frutos-Torres *et al.*, 2021). Por otro lado, para Giraldo-Luque y Fernández-Rovira (2020) se destaca una relación dicotómica entre la percepción del uso adictivo de dispositivos móviles y la necesidad de estar en estas plataformas, reflejando estrategias sofisticadas de control en la sociedad de la información. Esta dependencia de las redes sociales ha abierto la puerta a un nuevo fenómeno impulsado por la IA: los algoritmos de recomendación. Pariser (2012) acuñó el término «burbujas de filtro» para describir cómo estos algoritmos personalizan el contenido que los usuarios ven, limitando la exposición a puntos de vista diversos y contribuyendo a la creación de cámaras de eco. En este contexto, los usuarios jóvenes se ven expuestos a información que refuerza sus creencias, incrementando el riesgo de desinformación.

La IA ha intensificado este proceso. Plataformas como Google y Facebook emplean algoritmos avanzados que utilizan IA para optimizar los resultados de búsqueda y las recomendaciones de contenido en función del historial del usuario y sus interacciones previas (Tandoc *et al.*, 2018). Si bien esta personalización mejora la experiencia de usuario, también presenta desafíos significativos en términos de veracidad de la información. Como Tandoc señala, los algoritmos tienden a priorizar el contenido que genera más interacciones, sin necesariamente evaluar su veracidad.

En este sentido, la IA ha transformado no solo el acceso, sino también la creación y distribución de contenido. El informe de la UNESCO (Miao *et al.*, 2020) advierte sobre la creciente presencia de noticias falsas y desinformación generadas por IA, lo que exacerba la necesidad de competencias críticas entre los jóvenes para poder evaluar la veracidad de la información que consumen. En el caso de los universitarios, quienes a menudo dependen de aplicaciones de IA como motores de búsqueda o asistentes virtuales para obtener información, esta dependencia puede

generar una sobreexposición a contenido sesgado o inexacto, agravando el fenómeno de la desinformación (McStay, 2017).

Este contexto plantea la necesidad de abordar la alfabetización mediática en la era de la inteligencia artificial, un campo que está ganando cada vez más atención entre los investigadores. Vega *et al.* (2024) señalan que, aunque los jóvenes son usuarios avanzados de tecnología, su capacidad para discernir entre información veraz y noticias falsas es limitada, debido en gran parte a la opacidad con la que operan los algoritmos de las plataformas digitales. Esto sugiere que, aunque la IA ha facilitado el acceso a la información, también ha amplificado los desafíos relacionados con la veracidad y la pluralidad informativa.

2.2. Inteligencia artificial y desinformación

La IA ha transformado la difusión de la información y, simultáneamente, ha amplificado los riesgos relacionados con la desinformación. Los algoritmos de las redes sociales y motores de búsqueda son capaces de priorizar contenido sensacionalista o falso debido a su capacidad de generar mayor interacción, una problemática evidenciada por Vosoughi *et al.* (2018) en su estudio sobre la propagación de noticias falsas en Twitter. Los autores señalan que las noticias falsas tienen un 70% más de probabilidades de ser compartidas que las verdaderas, lo que demuestra el impacto del diseño algorítmico en la diseminación de desinformación. Como señalan Gómez-Calderón *et al.* (2021), los jóvenes actuales tienen una alta exposición a *fake news*, con variaciones significativas basadas en características sociodemográficas, lo que denota la importancia de la educación en alfabetización mediática y crítica para mitigar los efectos de la desinformación en este grupo etario. Para Gómez-Calderón *et al.* (2023), la capacidad de identificar noticias falsas aumenta con la edad y el nivel educativo. Además, quienes residen en grandes ciudades tienden a verificar la información con más frecuencia.

Por otro lado, Buchanan (2020) advierte que los sistemas de IA, como GPT-3, pueden ser utilizados para generar noticias falsas de manera masiva, simulando artículos de apariencia verídica. Esta capacidad ha puesto a prueba la alfabetización mediática de los usuarios, quienes se ven cada vez más expuestos a contenido difícil de distinguir de las fuentes legítimas. Según Helberger *et al.* (2016), los algoritmos de personalización crean burbujas informativas que limitan el acceso a perspectivas diversas, exacerbando los sesgos de confirmación entre los usuarios.

La IA también ha permitido el desarrollo de las *deepfakes*, tecnología que utiliza redes neuronales para crear videos falsos de personas. Chesney y Citron (2019) destacan que las *deepfakes* representan una amenaza grave para la integridad de la información visual, particularmente

en contextos políticos y sociales, donde su uso puede manipular opiniones y erosionar la confianza en los medios tradicionales.

Un análisis de Shu *et al.* (2020) identifica que las plataformas de IA facilitan tanto la creación como la propagación de noticias falsas, destacando la necesidad de herramientas más sofisticadas para detectar y eliminar contenido engañoso. Sin embargo, Ferrara (2017) subraya que los sistemas actuales de moderación de contenido aún no están suficientemente preparados para combatir la velocidad a la que se propaga la desinformación en redes como Facebook y Twitter, lo que agrava la magnitud del problema.

La intervención regulatoria también ha sido objeto de debate. La Comisión Europea (2024) ha identificado la necesidad de marcos normativos más estrictos para controlar el uso de IA en la difusión de información, implementando códigos de conducta para que las plataformas tecnológicas adopten prácticas más responsables. Sin embargo, Floridi (2019) sostiene que la regulación debe ir acompañada de una alfabetización digital más robusta que permita a los usuarios identificar noticias falsas de manera efectiva.

Finalmente, el trabajo de Howard y Neudert (2017) muestra cómo los actores políticos han utilizado bots y algoritmos de IA para influir en campañas electorales, difundir desinformación y manipular a los votantes a través de plataformas digitales. Estos estudios sugieren que, si bien la IA ha revolucionado el acceso a la información, también ha introducido nuevos desafíos éticos y regulatorios que deben ser abordados de manera urgente.

3 | Metodología

Partimos de la hipótesis de que los jóvenes andaluces utilizan, de algún modo, la IA para obtener información, bien sobre asuntos que le interesan o bien para la elaboración de algún trabajo de clase. En este sentido, es posible que este segmento de la población no contraste la información que les devuelve la IA, generando así un alto grado de desinformación o, incluso, infoxicación, término del que han hablado diferentes autores referida primero a los medios de comunicación (Scolari, 2015), después a la que llega a través de la red (Benito-Ruiz, 2009; Cornellá, 2010; Dias, 2014), pasando por la que se produce debido a la proliferación de las redes sociales (Portugal y Aguaded, 2020) y centrada hoy en los usos de la IA.

El principal objetivo del estudio que se presenta radica en conocer si los jóvenes de entre 18 y 20 años utilizan la IA e identificar el grado de desinformación que puede experimentar según el uso que hacen de este tipo de aplicaciones para obtener información.

3.1. Método de investigación

La encuesta ha sido el método de trabajo utilizado en esta investigación, ya que «permite obtener información acerca de una muestra de la población total» (Dorado, 2002). Además de ser «una técnica que depende del contacto directo o indirecto con todas las personas, cuyas actitudes, conductas o características son significativas para una determinada investigación» (Campbell y Katona, 1979). En este caso, se ha encuestado a alumnado de segundo de Periodismo de la Universidad de Sevilla, a quienes, siguiendo los pasos de Sierra Bravo (1988, p. 76), se les ha explicado de manera clara los objetivos de esta, se les ha respondido a las dudas que hayan podido tener de forma clara, se les ha dado la oportunidad de decidir si participar o no y los datos obtenidos se usan única y exclusivamente para los fines que se describían en el encabezado del cuestionario.

Como soporte para la realización de la encuesta se ha utilizado la aplicación de Google Form, con este encabezado:

Esta encuesta es totalmente anónima. La misma, tiene como único objetivo conocer dónde buscan información los jóvenes andaluces, su relación con la inteligencia artificial a la hora de buscar información y el grado de información/desinformación que pueda desprenderse de los resultados. Estos datos serán publicados en la investigación que lleva por título: Inteligencia artificial y desinformación. El caso de los jóvenes andaluces.

Las preguntas del cuestionario fueron:

- Edad
- Habitualmente, ¿a través de qué canales te informas?
- Si es en redes sociales, ¿cuál o cuáles?
- ¿Crees que te llega la información que buscas o la que el algoritmo te hace llegar?
- ¿Buscas la información o lees aquella que te llega?
- ¿Sueles ayudarte de alguna aplicación de IA para obtener información?
- ¿Has buscado alguna vez información en una aplicación de AI?
- ¿Cuál o cuáles aplicaciones usas o has usado?
- ¿Qué tipo de información has buscado?
- ¿Te ha resultado útil la información que te ha devuelto?
- ¿Has contrastado la información que te ha devuelto la IA?
- En caso de haber respondido cualquier opción excepto nunca, ¿había incorrecciones en la información?

Con los resultados se ha realizado un análisis cuantitativo y cualitativo y se han extraído conclusiones sobre los usos de los jóvenes de las

redes sociales, la IA y las repercusiones que pueden tener estos hábitos en el grado de información/desinformación de este sector poblacional.

4 | Discusión y resultados

El cuestionario propuesto se le pasó al alumnado de segundo curso del Grado en Periodismo de la Universidad de Sevilla y respondieron 159 personas. Del total de personas encuestadas, el 86,8% tenían una edad entre los 18 y los 20 años.

Resulta importante destacar que el 62,9% de los jóvenes encuestados utilizan para informarse, habitualmente, las redes sociales. La red social más utilizada para ello es X (Twitter) con el 47,8%, seguida por Tiktok, con la mitad de porcentaje que la primera (22,6%), muy igualada con Instagram, que la usa el 22% (fig. 4.1).

Pero quizá lo más alarmante es el hecho de que más de la mitad de estos jóvenes (57,9%) cree que le llega la información que aleatoriamente le hace llegar el algoritmo a sus perfiles y solo el 25,2% busca información para leer, el 65,4% asegura que lee tanto la información que busca como la que le llega aleatoriamente y solo el 9,4% asegura consumir la información que le llega sin buscarla. Un hecho que nos hace pensar que, realmente, consumen la información que les llega, porque resulta contradictorio que a la mayoría le llegue la información de forma aleatoria y que luego no sea esta la información que lee.

Más allá del uso de las redes sociales, interesa en este estudio conocer los hábitos de consumo de inteligencia artificial por parte de los jóvenes andaluces. En este sentido, más de la mitad se ha ayudado alguna vez de alguna aplicación de IA para obtener información y el 85,5% ha buscado alguna vez información en una de ellas, con lo

Si es en redes sociales, ¿Cuál o cuáles?

159 respuestas

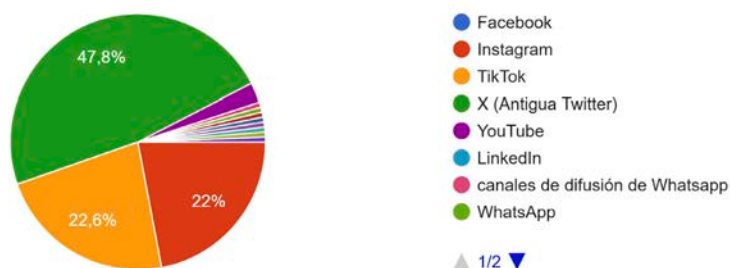


Figura 4.1. Redes sociales a través de las que se informan los jóvenes andaluces.

Fuente: elaboración propia

¿Has buscado alguna vez información en una aplicación de AI?

159 respuestas

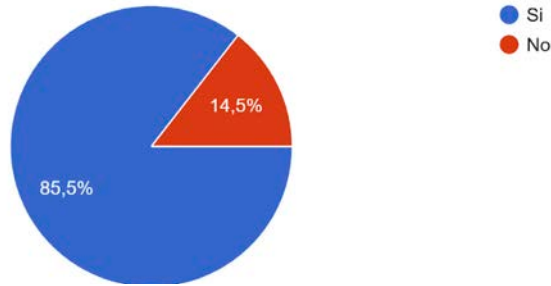


Figura 4.2. Jóvenes andaluces que han buscado información en aplicaciones IA. Fuente: elaboración propia

¿Has contrastado la información que te ha devuelto la IA?

159 respuestas

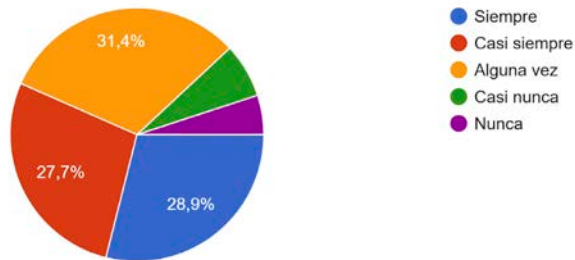


Figura 4.3. Grado de contraste de información obtenida a través de la IA. Fuente: elaboración propia

cual, nos podemos quedar con este segundo dato, afirmando que la gran mayoría de los encuestados ha usado IA para obtener información (fig. 4.2.).

Respecto a la aplicación más utilizada, arrasa ChatGPT, el 91,2% dice haberla usado. Resulta llamativo que 145 personas aseguren haber usado esta aplicación, puesto que algunos dijeron antes que no habían usado la IA para obtener información. El 78% de ellos usaron la IA para algún trabajo de clase y solo el 10,7% la usaron para buscar noticias. Afirman, además, que al 73,42% le ha resultado útil, siempre o casi siempre, la información que le ha devuelto la aplicación, solo el 5,6% dice que no le fue útil nunca o casi nunca.

Aunque, lo más alarmante que se desprende de este estudio, es que solo el 56,6% ha contrastado la información siempre, o casi siempre, el 31,4% solo alguna vez y el 11,9% no lo ha hecho nunca, o casi nunca (fig. 4.3).

De las personas que han contrastado la información, el 56,6% asegura que siempre, casi siempre o alguna vez han encontrado incorrecciones, mientras que el 43,3% no ha encontrado incorrecciones nunca o casi nunca. Curiosamente, este dato se corresponde con el 43,3% que solo alguna vez o nunca ha contrastado la información, aunque no se puede establecer un paralelismo inmediato, sí podemos decir que de las 69 personas que solo alguna vez, nunca o casi nunca han contrastado la información, 38 nunca han encontrado incorrecciones, lo cual puede resultar lógico.

Conclusiones

Tras el análisis de los datos obtenidos, la primera conclusión radica en que casi el noventa por ciento de los jóvenes andaluces encuestados ha usado la IA en algún momento para buscar información y solo el 56,6% la ha contrastado, siempre o casi siempre. Por lo tanto, el 43,3% de los jóvenes andaluces puede haber recibido información falsa sin contrastar, si no siempre, sí en alguna ocasión. Un hecho que debe hacernos reflexionar a quienes trabajamos con jóvenes sobre la falta de información acerca de las incorrecciones que pueden derivarse de los datos obtenidos por aplicaciones de IA y la importancia que esto puede tener, ya no solo en la desinformación por parte de la juventud andaluza, sino en lo fácilmente que pueden ser receptores de informaciones interesadas, sin saberlo.

Los jóvenes andaluces consumen la información que les llega a través de las redes sociales sin pedirla, es decir, es el algoritmo quien decide, en buena medida, la información que consumen. Además, buscan información en aplicaciones de IA, sobre todo en ChatGPT, no siempre contrastando los resultados de la búsqueda con otras fuentes. Estos dos hechos derivan, necesariamente, en dos conclusiones, la primera es que los jóvenes andaluces encuestados pueden ser fácilmente manipulables por parte de quienes controlan los algoritmos que gestionan la información y, por otro, pueden sufrir una gran desinformación, no solo porque no les llega toda la información a través de las plataformas que usan para recabarla, sino porque ni siquiera siempre contrastan aquella que ellos mismos buscan a aplicaciones de inteligencia artificial. Todo ello, puede derivar en una infoxicación importante, creen tener grandes cantidades de información, no siempre verídica, que les cuesta gestionar y contrastar.

Referencias

- Benito-Ruiz, E. (2009). "Infocination 2.0". En Thomas, M. (ed.): *Handbook of Research on Web 2.0 and Second Language Learning* (60-79). IGI Global. <https://doi.org/10.4018/978-1-60566-190-2.ch004>.
- Buchanan, B. (2020). *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics*. Harvard University Press.
- Chesney, R. y Citron, D. (2019). Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. *Foreign Affairs*, 98(1), 147-155. <https://www.jstor.org/stable/26798018>.
- Comisión Europea (2024). Lucha contra la desinformación en línea. <https://bit.ly/42gpPTe>.
- Cornellá, A. (2004). *Infocixación: Buscando un orden en la información*. Infomanía. <https://bit.ly/40CTaGq>.
- De Frutos-Torres, B., Pastor-Rodríguez, A. y Cruz-Díaz, R. (2021). Credibilidad e implicaciones éticas de las redes sociales para los jóvenes. *Revista Latina de Comunicación Social*, (79), 51-68. <https://doi.org/10.4185/RLCS-2021-1512>.
- Dias, P. (2014). From 'infocination' to 'infosaturation': a theoretical overview of the cognitive and social effects of digital immersion. *Ámbitos: Revista Internacional de Comunicación*. 24, 31-40. <http://dx.doi.org/10.12795/Ambitos.2014.i24.04>.
- Ferrara, E. (2017). Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election. *First Monday*, 22(8). <https://doi.org/10.5210/fm.v22i8.8005>.
- Floridi, L. (2019). *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford Academy. <https://doi.org/10.1093/oso/9780198833635.001.0001>.
- Giraldo-Luque, S. y Fernández-Rovira, C. (2020). Redes sociales y consumo digital en jóvenes universitarios: economía de la atención y oligopolios de la comunicación en el siglo XXI. *Profesional De La Información*, 29(5). <https://doi.org/10.3145/epi.2020.sep.28>.
- Gómez-Calderón, B., Córdoba-Cabús, A. y Méndez-Nieto, A. (2021). Jóvenes y fake news: Un análisis sociodemográfico aplicado al caso andaluz. *IC Revista Científica de Información y Comunicación*, (17). <https://bit.ly/4gW9RlD>.
- Gómez-Calderón, B., Córdoba-Cabús, A. y López-Martín, Á. (2023). Las fake news y su percepción por parte de los jóvenes españoles: El influjo de los factores sociodemográficos. *Doxa Comunicación. Revista Interdisciplinaria De Estudios De Comunicación Y Ciencias Sociales*, 36, 19-42. <https://doi.org/10.31921/doxacom.n36a1741>.
- Helberger, N., Karppinen, K. y D'Acunto, L. (2016). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2), 191-207. <https://doi.org/10.1080/1369118X.2016.1271900>.
- Howard, P. N. y Neudert, L. M. (2017). *Computational propaganda in the United States of America: Manufacturing consensus online*. University of Oxford. <https://bit.ly/3PC5KPR>.

- Kantar (2021). *Solo uno de cada tres jóvenes de entre 16 y 24 años sabe identificar la fuente de la información*. Kantar. <https://bit.ly/3WjCSV1>.
- McStay, A. (2017). *Privacy and the Media*. SAGE Publications.
- Miao, F., Holmes, W., Huang, R. y Zhang, H. (2020). *AI and education: Guidance for policy-makers*. United Nations Educational, Scientific and Cultural Organization. <https://doi.org/10.54675/PCSP7350>.
- Pariser, E. (2012). *The filter bubble: What the internet is hiding from you*. Penguin Books Ltd.
- Portugal, R. y Aguaded, I. (2020). Competencias mediáticas y digitales, frente a la desinformación e infoxicación. *Razón y Palabra*, 24(108), 5-36. <https://doi.org/10.26807/RP.V24I108.1658>.
- Scolari, C. A. (2015). *Ecología de los medios. Entornos, evoluciones e interpretaciones*. Gedisa Editorial.
- Shu, K., Wang, S., Lee, D. y Liu, H. (2020). Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements. En: Shu, K., Wang, S., Lee, D. y Liu, H. (eds.): *Disinformation, Misinformation, and Fake News in Social Media. Lecture Notes in Social Networks*. Springer, Cham. https://doi.org/10.1007/978-3-030-42699-6_1.
- Tandoc, E. C., Lim, Z. W. y Ling, R. (2017). Defining "Fake News": A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153. <https://doi.org/10.1080/21670811.2017.1360143>.
- Vega Cocha, D. P., Vásquez Chicaiza, F. P., Guevara Guevara, A. M., Maldonado Palacios, I. A. y Maldonado Arce, M. E. (2024). Desinformación en la era digital: el papel de las redes sociales en la propagación de noticias falsas durante conflictos globales. *Revista Latinoamericana de Ciencias Sociales y Humanidades*, V(2), 138-149. <https://doi.org/10.56712/latam.v5i2.1865>.
- Vosoughi, S., Roy, D. y Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>.

Retos y oportunidades de la inteligencia artificial: uso y percepción entre adolescentes, familias y docentes

Xavier Bringué Sala

Universidad de Navarra
jbringue@unav.es

Charo Sádaba Chalezquer

Universidad de Navarra
csadaba@unav.es

1 | Introducción

El avance de la inteligencia artificial generativa (IAG) ha abierto nuevas posibilidades en diversos campos, incluido el ámbito educativo. Herramientas como ChatGPT, DALL-E y otras plataformas de generación de contenido están transformando la manera en que los estudiantes acceden a la información, resuelven problemas y desarrollan habilidades creativas. Sin embargo, este impacto genera una creciente preocupación entre las familias y los docentes, quienes cuestionan cómo estas tecnologías pueden influir en los hábitos de aprendizaje, la capacidad crítica y la autonomía de los adolescentes en su formación académica. También hay preocupación por las implicaciones éticas que conlleva su uso y cómo se trasladan al ámbito educativo.

Uno de los potenciales beneficios que se señala de la IAG en la educación es su capacidad para personalizar el aprendizaje. Los estudiantes pueden recibir explicaciones detalladas, ejemplos prácticos y soluciones adaptadas a su nivel de comprensión y estas ser repetidas tantas veces como sea preciso sin interferir en el curso del resto de la clase. Además, estas herramientas pueden ayudar en la redacción de textos, la resolución de problemas matemáticos y la generación de ideas para proyectos, lo que permite un acceso más eficiente al conocimiento. Sin embargo, esta facilidad de acceso va de la mano de riesgos como la disminución del esfuerzo cognitivo, la dependencia excesiva de la tecnología y la posible pérdida de habilidades fundamentales como la escritura y la argumentación crítica.

Los equipos docentes han expresado su preocupación respecto a la dificultad de evaluar el trabajo de los estudiantes en un contexto donde la IAG puede generar respuestas completas y bien estructuradas en segundos. Esto plantea desafíos en la detección del plagio, la autenticidad del pensamiento y la valoración del proceso de aprendizaje en lugar del simple resultado final. Además, existe el riesgo de que los adolescentes no desarrollen una adecuada alfabetización digital, lo que les impida discernir entre información veraz y contenido generado por IA que podría contener errores o sesgos y que estarían validando y asumiendo como propio sin ser conscientes de las implicaciones que esto puede tener en su imagen personal.

Por su parte, las familias también muestran inquietudes sobre el impacto que el uso indiscriminado de la IAG puede tener en la autonomía intelectual y la capacidad de resolución de problemas de sus hijos. La facilidad con la que los estudiantes pueden obtener respuestas inmediatas podría desincentivar la exploración, la curiosidad y el pensamiento crítico, elementos esenciales para un aprendizaje significativo. Además, preocupa el impacto emocional y psicológico que estas herramientas pueden tener en los adolescentes, especialmente en términos de autoeficacia y confianza en sus propias capacidades intelectuales.

Junto con esto, es evidente que se ha instalado un cierto discurso social favorable a la adopción de la IA de una manera rápida y aparentemente inevitable, que lleva consigo el requerimiento de una reforma educativa drástica que capacite a los estudiantes para hacer frente a esta nueva realidad. La idea de un futuro inmediato donde se va a requerir a quienes acceden al mercado laboral competencias y capacidades relacionadas con la IAG que todavía no se conocen genera un cierto temor. No obstante, este discurso también se ha visto enfrentado por otros que reclaman una moratoria en el desarrollo de la IAG para poder reflexionar en sus implicaciones y que permita un ritmo más humano de desarrollo. Aunque estas reclamaciones no han sido atendidas, al menos han servido para poner de manifiesto que, aunque sí pueda percibirse un discurso más popular a favor de la IAG, no se trata de un discurso único.

Ante estos desafíos, parece fundamental promover un uso equilibrado y reflexivo de la inteligencia artificial generativa en la educación. Esto implica la implementación de estrategias pedagógicas que integren la IAG como una herramienta de apoyo, pero sin reemplazar la enseñanza tradicional ni la necesidad de que los estudiantes desarrollen habilidades analíticas y creativas. Asimismo, es necesario fortalecer la alfabetización digital para que los adolescentes aprendan a utilizar estas tecnologías de manera crítica y responsable.

Aunque todo apunta a que la inteligencia artificial generativa representa una oportunidad sin precedentes para mejorar el aprendizaje de los adolescentes, también plantea desafíos significativos que requieren una respuesta coordinada por parte de docentes, familias e instituciones educativas. Solo a través de una integración consciente y regulada de estas herramientas se podrá garantizar que su uso potencie el aprendizaje sin comprometer el desarrollo integral de los jóvenes.

En cualquier caso, parece preciso que las eventuales medidas que se puedan implementar partan de investigación que aporte evidencias sobre las que diseñar acciones y posibles líneas de actuación. Este es el ánimo de la investigación que da lugar a este texto, así como esta aportación en concreto. Se trataba de conocer el nivel de uso que adolescentes, familias y profesores hacen de herramientas de IAG, sus preocupaciones y los retos que plantea, desde su punto de vista. Con este fin, Empantallados, una plataforma para padres y madres que busca ofrecer recursos formativos a las familias para que puedan acompañar a sus hijos e hijas también en el mundo digital, dedicó su estudio anual en el año 2023 a conocer «El impacto de la IA en la educación en España». El trabajo de campo fue llevado a cabo por la empresa de investigación de mercados GAD3 y la base de datos obtenida se puso a disposición de los investigadores autores de este capítulo para que pudieran dar difusión académica a sus resultados.

2 | Revisión de la literatura

Aunque la investigación sobre inteligencia artificial no es nueva y ha sido habitual en las últimas décadas, la inteligencia artificial generativa (IAG) es un avance más reciente que ha marcado una disrupción en la evolución de la IA. Esto hace que la investigación sobre el impacto de la IAG sea todavía incipiente y, más todavía, la relacionada con su impacto en la educación. No obstante, la rápida adopción de algunas herramientas de IAG por parte de la población estudiantil ha levantado las alarmas y han comenzado a aparecer estudios que, pese a ser incipientes, señalan ámbitos claros de interés, como el impacto en el proceso de aprendizaje, su capacidad para personalizar la enseñanza y los desafíos éticos que conlleva su implementación.

El trabajo de Chassignol *et al.* (2018) repasaba las narrativas más habituales asociadas al uso y al impacto de la IA en la educación en varios estudios académicos. De acuerdo con su estudio, se percibía que, pese a estar cambiando gracias al uso de algunas herramientas de AI, no parece pensarse que estemos ante un reemplazo del sistema educativo tradicional. La idea de que la IA, por ejemplo, pueda eliminar la necesidad de interacción social propia del entorno educativo, no parece estar contemplada. Se consolidaba la idea de más bien se trata de añadirla al proceso de aprendizaje tradicional como ya se ha hecho con la gaendizaje de los más pequeños, manifestaron preocupaciones sobre un tiempo de pantalla excesivo y la falta de actividad física y de interacción social que su uso puede suponer, refrendando cuestiones que vienen siendo habituales (Ahearne *et al.*, 2016; Blackwell *et al.*, 2014; Wood *et al.*, 2008). El estudio de Quan *et al.* (2024) pone de manifiesto, sin embargo, que las actitudes positivas de padres y madres hacia ChatGPT y su apoyo de esta herramienta están significativamente asociados al uso que ellos y ellas hagan de ChatGPT, que influye también de manera significativa, el de sus hijos (Otermans *et al.*, 2024).

Estas investigaciones previas refuerzan la importancia de un enfoque equilibrado en la implementación de la IA en el ámbito educativo. Si bien esta tecnología ofrece numerosas oportunidades para mejorar el aprendizaje, también requiere regulaciones adecuadas y un uso consciente para maximizar sus beneficios sin comprometer la formación integral de los estudiantes. También se pone de manifiesto que hay mucho que seguir investigando sobre el tema en múltiples perspectivas antes de tomar decisiones que puedan tener efectos no deseados o nocivos.

3 | Metodología

Con el fin de recoger información sobre los usos, las percepciones y las preocupaciones que genera el uso de la IA entre los menores, se planteó realizar una encuesta a población residente en España con hijos menores de 18 años, adolescentes entre 14 y 18 años y profesores de colegios.

La muestra consistió en 901 personas que representaban por variable de sexo y edad a la población española. 501 de los encuestados eran padres o madres, 200 adolescentes y 200 profesores. La entrevista se realizó online a través de un cuestionario autoadministrado. La empresa de investigación de mercados GAD3, encargada de la recogida de datos, fue también la responsable de la gestión de los consentimientos de todos los participantes.

La duración media de la entrevista, de 25 preguntas, fue de entre 7-8 minutos. La recogida de información tuvo lugar entre octubre y diciembre de 2023. Los datos arrojan un margen de error del $\pm 3\%$ ($n=901$) sobre la muestra total y del $\pm 7,1\%$ ($n=200$) para los grupos más pequeños para un

grado de confianza del 95,5% (dos sigmas) y en la hipótesis más desfavorable de $P=Q=0,5$ en el supuesto de muestreo aleatorio simple.

La investigación se fundamenta en un enfoque descriptivo y explicativo, permitiendo analizar la adopción y el impacto de herramientas de IA desde múltiples perspectivas. La recolección de datos se realizó a través de cuestionarios aplicados a tres grupos clave: estudiantes, padres y docentes. Estas encuestas incluyen preguntas cerradas y escalas de valoración que permiten medir el nivel de uso, confianza y expectativas en torno a la IA.

Además, el estudio incorpora un análisis de la percepción sobre el impacto educativo de la IA, evaluando cómo estas herramientas influyen en el desempeño académico y en los procesos de enseñanza. También se examinan las preocupaciones sobre la privacidad y la fiabilidad de la información proporcionada por la IA.

Como se ha dicho, la muestra del estudio abarca una representación significativa de adolescentes, padres y docentes, seleccionados de manera que reflejen diferentes niveles educativos y contextos socioculturales. Se utilizó un muestreo intencional para garantizar la diversidad en el uso y conocimiento de la IA dentro del ámbito educativo.

Cada grupo respondió a una serie de preguntas adaptadas a su experiencia con la IA. Por ejemplo, los estudiantes detallaron las herramientas que han utilizado y sus usos específicos, mientras que los docentes indicaron cómo han integrado la IA en su enseñanza. Los padres, por su parte, compartieron su percepción sobre la utilidad y los riesgos asociados a estas tecnologías.

3.1. Instrumentos de medición

El cuestionario incluyó diferentes tipos de preguntas ($n=25$) para abordar diversos aspectos del uso y la percepción de la IA. Algunas de las preguntas más relevantes fueron:

- **Uso de herramientas de IA:** identificación de herramientas utilizadas, frecuencia de uso y propósito.
- **Conocimiento sobre IA:** canales de información sobre IA y nivel de comprensión.
- **Percepción sobre la fiabilidad:** evaluación de la objetividad de la IA y sus posibles sesgos.
- **Impacto educativo:** opinión sobre cómo la IA afecta el aprendizaje y la enseñanza.
- **Retos y necesidades:** expectativas sobre la formación en IA y preocupaciones éticas.

4 | **Discusión y resultados**

4.1. **Conocimiento y uso de herramientas de IA**

Los resultados revelan que el uso de herramientas de inteligencia artificial generativa (IAG) está ampliamente extendido entre los adolescentes y los profesores. Un 82 % de los estudiantes y un 73 % de los docentes han utilizado herramientas como ChatGPT o asistentes virtuales, lo que indica una rápida integración de estas tecnologías en el entorno educativo. Este alto porcentaje refleja no solo el acceso creciente a estas herramientas, sino también su utilidad en diversos ámbitos académicos.

Los docentes han incorporado la IAG principalmente para la generación de materiales educativos, la planificación de clases y la evaluación del desempeño estudiantil. Por su parte, los estudiantes han manifestado utilizar estas herramientas para tareas como la redacción de textos, la resolución de problemas matemáticos y la búsqueda de información académica. No obstante, un porcentaje significativo de alumnos (56 %) reconoce que la IA puede proporcionar información inexacta o sesgada, lo que sugiere una percepción más crítica sobre su fiabilidad.

En cuanto a los padres y madres, se observa una menor familiaridad con las herramientas de IAG, aunque un 60 % de ellos ha utilizado alguna forma de inteligencia artificial en su vida cotidiana, principalmente en aplicaciones de reconocimiento de voz o asistentes virtuales. Esta brecha en el uso sugiere la necesidad de programas de formación para que los padres puedan comprender mejor cómo estas tecnologías influyen en la educación de sus hijos.

4.2. **Percepciones y preocupaciones**

El impacto de la inteligencia artificial en la educación genera sentimientos encontrados entre familias y docentes. Mientras que el 57 % de los padres y madres otorgan una valoración positiva a la IA en el ámbito educativo, solo el 39 % de los docentes consideran que puede aportar ventajas significativas en la enseñanza. Esta diferencia sugiere una mayor reticencia por parte del personal educativo, posiblemente debido a la preocupación sobre la calidad del aprendizaje y la autenticidad del trabajo de los alumnos.

En términos de privacidad y protección de datos, las inquietudes son notables. Un 83 % de las familias y un 90 % de los docentes expresan preocupación por el manejo de los datos personales en las plataformas de IA, lo que refuerza la necesidad de establecer normativas claras y estrategias de formación en seguridad digital. Muchos padres manifiestan su

temor de que sus hijos compartan información sensible sin darse cuenta de las implicaciones que esto conlleva.

La objetividad y precisión de la IA también son puntos de controversia. Padres y madres califican la fiabilidad de la IA con un 6,2 sobre 10, mientras que los profesores le otorgan un 5,2. Esta percepción sugiere que, si bien existe cierto grado de confianza en la tecnología, aún hay dudas sobre la veracidad de la información generada. La preocupación por la propagación de sesgos y la posible manipulación de la información sigue siendo un tema recurrente en las entrevistas realizadas.

4.3. Impacto educativo

En relación con el impacto educativo, los resultados muestran una diversidad de opiniones. En una escala de 0 a 10, donde 0 indica un impacto negativo y 10 uno positivo, los participantes han evaluado la influencia de la IA en la educación de la siguiente manera.

Las **familias** consideran que la IA impactará de forma moderadamente positiva el desempeño de los alumnos, con una puntuación promedio de 6,5. Muchos padres y madres ven en la IA una herramienta que puede complementar el aprendizaje de sus hijos, permitiéndoles acceder a explicaciones detalladas y personalizadas. No obstante, algunos expresan preocupaciones sobre la posibilidad de que los estudiantes se vuelvan dependientes de estas tecnologías y reduzcan su esfuerzo en el aprendizaje autodidacta. Además, existe la inquietud de que la IA pueda reemplazar la interacción humana en la enseñanza, disminuyendo la capacidad de los alumnos para desarrollar habilidades sociales y colaborativas.

Los **docentes** muestran mayor escepticismo, con una calificación media de 5,8 en cuanto a la mejora del rendimiento estudiantil. A pesar de que algunos reconocen los beneficios de la IA en la generación de contenido y la optimización del tiempo de enseñanza, existe un temor generalizado de que los estudiantes la utilicen como un atajo para evitar el esfuerzo intelectual. Los profesores también resaltan la falta de regulación y estrategias claras para su implementación en el aula, lo que genera incertidumbre sobre su impacto a largo plazo.

Por su parte, el **estudiantado** tiene una percepción más favorable, con una valoración de 7,1, lo que sugiere que ven la IA como una herramienta útil para su aprendizaje. La mayoría de los alumnos consideran que la IA les permite acceder a información de manera rápida y eficiente, facilitando la comprensión de temas complejos. Sin embargo, algunos estudiantes también reconocen que la IA puede ser engañosa o proporcionar información inexacta, lo que subraya la importancia de desarrollar habilidades de verificación y análisis crítico en su uso.

Los docentes han señalado la necesidad de adaptar sus metodologías de enseñanza para integrar la IA de manera efectiva. Un 73% de los profesores han utilizado IA como un recurso de apoyo, pero insisten en que no debe reemplazar el papel fundamental del educador. En este sentido, los esfuerzos se centran en la combinación de herramientas digitales con estrategias pedagógicas tradicionales que fomenten el pensamiento crítico y la creatividad.

Además, los resultados sugieren que la IA puede desempeñar un papel importante en la personalización del aprendizaje. Algunos docentes han comenzado a utilizar estas herramientas para adaptar materiales a distintos niveles de competencia, permitiendo que los estudiantes avancen a su propio ritmo. Sin embargo, también existen preocupaciones sobre la equidad en el acceso a estas tecnologías, ya que no todos los alumnos disponen de los mismos recursos tecnológicos en sus hogares.

4.4. Retos y necesidades

Uno de los principales desafíos identificados es la falta de formación en el uso ético y seguro de la IA. Aunque un 82% de los docentes, el 78% de los padres y el 63% de los estudiantes desean aprender más sobre estas herramientas, existe una brecha en la alfabetización digital que impide su aprovechamiento óptimo.

La necesidad de desarrollar habilidades críticas en los estudiantes también es un tema clave. Los profesores enfatizan la importancia del pensamiento crítico para evaluar la veracidad de la información generada por la IA. En este sentido, un enfoque educativo centrado en la enseñanza del análisis y la evaluación de datos puede contribuir a un uso más responsable de la tecnología.

Por otro lado, el 61% de los padres y el 54% de los docentes consideran que la IA puede ser un recurso positivo para el futuro laboral de los estudiantes. Esto resalta la importancia de preparar a los jóvenes para un mercado laboral donde la inteligencia artificial jugará un rol central.

Algunos docentes han expresado su preocupación por el efecto que el uso excesivo de la IA puede tener en la capacidad de los estudiantes para desarrollar habilidades de resolución de problemas sin asistencia tecnológica. Esto ha llevado a algunos colegios a establecer regulaciones específicas sobre el uso de la IA en las aulas, limitando su empleo en ciertos contextos académicos.

Finalmente, la inclusión de la IA en el currículo académico es vista como una oportunidad, pero requiere una regulación clara para evitar un uso indebido. Un 68% de los docentes consideran que es necesario establecer directrices específicas sobre cómo y en qué contextos se deben utilizar estas herramientas en el aula. Se destaca la importancia

de equilibrar el acceso a la IA con la enseñanza de habilidades fundamentales como la escritura, la argumentación lógica y la creatividad

Conclusiones

Los hallazgos de esta investigación destacan la creciente presencia de la inteligencia artificial generativa en la educación y la diversidad de percepciones sobre su impacto. Si bien la mayoría de los estudiantes y docentes han adoptado estas herramientas, persisten preocupaciones sobre la privacidad, la fiabilidad y la influencia en el aprendizaje.

Para garantizar un uso adecuado de la IA en el entorno educativo, es esencial fomentar la formación en alfabetización digital, establecer regulaciones claras y promover metodologías pedagógicas que equilibren el uso de la tecnología con el desarrollo de habilidades fundamentales. La IA no debe ser vista como un sustituto del aprendizaje tradicional, sino como un complemento que potencie la capacidad crítica y creativa de los estudiantes.

El futuro de la IA en la educación dependerá en gran medida de cómo se integre en los sistemas de enseñanza y de la capacidad de las instituciones para establecer un marco ético y normativo adecuado. Con una implementación equilibrada y una formación adecuada, la inteligencia artificial puede convertirse en una herramienta poderosa para mejorar la calidad educativa y preparar a las nuevas generaciones para los desafíos del siglo XXI.

Recomendaciones

Son muchas las posibles recomendaciones que se siguen de este primer estudio, aunque quizá la más relevante es la necesidad de seguir investigando y prestando atención a la evolución del fenómeno y a sus usos entre los distintos públicos.

Las **administraciones públicas** tienen la oportunidad de trabajar en la implementación de programas de formación en inteligencia artificial (IA) dirigidos a docentes, estudiantes y familias con el propósito de garantizar un uso responsable y crítico de estas herramientas. Es fundamental desarrollar normativas claras sobre la privacidad y seguridad de los datos en el uso de la IA en el ámbito educativo, así como promover la equidad en el acceso a estas tecnologías, asegurando que todos los estudiantes cuenten con los recursos necesarios para aprovechar su potencial. Además, la creación de marcos regulatorios que definan con precisión los límites éticos y educativos en la integración de la IA en el currículo académico resulta esencial.

Desde la perspectiva de la **investigación**, se requiere un análisis exhaustivo sobre el impacto de la IA en el aprendizaje, evaluando tanto sus beneficios como sus riesgos. Es necesario estudiar cómo la IA puede contribuir a la personalización de la educación sin comprometer el desarrollo del pensamiento crítico y la autonomía del estudiante. Asimismo, la investigación debe abordar estrategias pedagógicas innovadoras que permitan la integración de la IA sin generar una dependencia tecnológica excesiva. Un aspecto clave en este ámbito es la evaluación de las diferencias socioeconómicas en el acceso a la IA y su impacto en la brecha educativa.

En el contexto **familiar**, resulta crucial fomentar el desarrollo del pensamiento crítico en niños y adolescentes, de modo que sean capaces de cuestionar y evaluar la información generada por la IA. La supervisión del uso de estas herramientas en el hogar es una medida necesaria para establecer límites adecuados y promover su empleo como un apoyo educativo, sin que sustituya el esfuerzo personal. Además, las familias deben informarse sobre las políticas de privacidad y protección de datos de las plataformas de IA utilizadas por sus hijos, así como establecer un diálogo abierto con los docentes y las instituciones educativas sobre su implementación en el aprendizaje.

El presente estudio presenta diversas fortalezas y limitaciones. Entre sus principales fortalezas, destaca la amplitud de la muestra, que incluye estudiantes, docentes y padres, lo que permite una visión integral sobre el impacto de la IA en la educación. Asimismo, el enfoque mixto, basado en la combinación de encuestas cuantitativas con entrevistas cualitativas, proporciona una comprensión más profunda de las percepciones y experiencias de los participantes. Finalmente, la actualidad del estudio en un contexto de rápida evolución tecnológica ofrece información relevante para la toma de decisiones educativas y regulatorias. No obstante, existen ciertas limitaciones. La representatividad de la muestra podría estar sesgada hacia instituciones con mayor acceso a tecnologías de IA, lo que restringe la generalización de los resultados. Además, la evolución constante de estas herramientas implica que algunos hallazgos podrían volverse obsoletos en un corto período de tiempo. Por último, la posible falta de conocimiento previo de los encuestados sobre la IA puede haber influido en sus respuestas, condicionando la percepción de sus impactos y limitaciones.

Referencias

- Ahearne, C., Dilworth, S., Rollings, R., Livingstone, V. y Murray, D. (2016). Touch-screen technology usage in toddlers. *Archives of Disease in Childhood*, 101(2), pp. 181-183.
- Akgun, S. y Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI Ethics* 2, 431-440. <https://doi.org/10.1007/s43681-021-00096-7>.
- Aslan, S., Durham, L. M., Alyuz, N., Chierichetti, R., Denman, P. A., Okur, E., y Nachman, L. (2024). What is the impact of a multi-modal pedagogical conversational AI system on parents' concerns about technology use by young children? *British Journal of Educational Technology*, 55(4), 1625-1650. <https://doi.org/10.1111/bjet.13399>.
- Blackwell, C. K., Lauricella, A. R., y Wartella, E. (2014). Factors influencing digital technology use in early childhood education. *Computers & Education*, 77, 82-90.
- Chassignol, M., Khoroshavin, A., Klimova, A., y Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia computer science*, 136, 16-24. <https://doi.org/10.1016/j.procs.2018.08.233>.
- Empantallados, GAD3 (2023). *El impacto de la IA en la educación en España. Familias y escuelas ante la Inteligencia Artificial*. <https://empantallados.com/ia/>.
- Otermans, P. C. J., Baines, S., Pereira, M., Livingstone, C. y Aditya, D. (2024). Chatting with the Future: A Comprehensive Exploration of Parents' Perspectives on Conversational AI Implementation in Children's Education. *International Journal of Technology in Education*, 7(3), 573-586. <https://doi.org/10.46328/ijte.812>.
- Quan, S., Du, Y. y Ding, Y. (2024, May). Young Children and ChatGPT: Parents' Use of ChatGPT in Parenting. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (1-7). <https://doi.org/10.1145/3613905.3650880>.
- Wood, E., Specht, J., Willoughby, T. y Mueller, J. (2008). Integrating computer technology in early childhood education environments: Issues raised by early childhood educators. *The Alberta Journal of Educational Research*, 54(2), 210-226.

El sexismo digital y la inteligencia artificial, otras formas de desigualdad y exclusión

Lucía Benítez Eyzaguirre

Universidad de Cádiz

lucia.benitez@uca.es

1 | Introducción

El sexismo digital describe cómo la tecnología y la programación refuerzan y ocultan formas de discriminación de género. Estos sesgos impactan negativamente en el acceso y uso de productos digitales para mujeres y minorías, sin que haya una conciencia pública o personal significativa que los cuestione. La digitalización no solo reproduce, sino que amplifica las desigualdades de género, afectando las actitudes y prácticas discriminatorias en plataformas tecnológicas. En este contexto, la tecnología ha sido históricamente construida bajo valores masculinos, como la racionalidad, la competencia y el control, lo que ha excluido a las mujeres, tanto de la influencia tecnológica como de los avances que representa. El estándar masculino ha dominado el desarrollo tecnológico, condicionando las experiencias digitales y la visibilidad de las mujeres.

Este análisis se basa en una revisión de la literatura académica y observaciones socioculturales sobre el uso de la tecnología, apoyado en una autoetnografía feminista. Este enfoque permite identificar valores culturales que asocian la tecnología con la guerra y el dominio, relegando

sus posibles usos para el bienestar humano y el cuidado. El estudio examina cómo estos valores han formado un modelo de desarrollo digital que reproduce estructuras de poder masculinas, desde la cultura del «brogrammer» hasta las representaciones sociales de las mujeres en roles subordinados, mientras los hombres son retratados como innovadores tecnológicos. También se analizan las implicaciones de esta cultura en la inteligencia artificial y la representación digital de las mujeres.

El estudio muestra patrones recurrentes de cosificación de las mujeres en entornos digitales, desde videojuegos y productos de robótica hasta asistentes virtuales. Estos productos tienden a representarlas de manera subordinada, reflejando un imaginario tecnológico que posiciona a los hombres como creadores y prioriza la competitividad, lo que a su vez excluye a las mujeres y la diversidad. La mercantilización de la cultura masculina y su enfoque en el control y la competencia ha segmentado el mercado laboral tecnológico, dificultando el acceso de las mujeres y limitando su capacidad para influir en la dirección de las innovaciones tecnológicas.

El diseño de las estructuras de seguridad y privacidad en los productos digitales también refleja una perspectiva masculina. Esto ha facilitado un entorno hostil para las mujeres en las plataformas digitales, consolidando la manófera como un espacio donde proliferan el acoso, la desinformación y la violencia simbólica, donde las mujeres huyen de debates digitales por temor a represalias y campañas de acoso. Además, la digitalización de estereotipos de género refuerza lógicas discriminatorias y su impacto en las interacciones *online*, en interfaces y aplicaciones, como los filtros de belleza.

La cultura tecnológica dominante, construida bajo un modelo masculino, perpetúa la exclusión de las mujeres y refuerza estereotipos de género a través de la inteligencia artificial y los algoritmos sesgados. Es necesario adoptar enfoques más inclusivos y equitativos en el diseño y desarrollo tecnológico para garantizar que las mujeres puedan participar plenamente y tener influencia en la creación de nuevas tecnologías que reflejen una diversidad de experiencias y necesidades.

2 | Revisión de la literatura

El sexismo digital no se limita a las actitudes y prácticas sexistas en el mundo digital ni a sus manifestaciones en las plataformas. Se refiere a la forma en que la propia tecnología reproduce y amplifica la desigualdad de género, un fenómeno de gran impacto que suele pasar desapercibido tanto para las personas usuarias como para la opinión pública y los actores políticos y sociales.

En este contexto, el desarrollo de la inteligencia artificial ha estado marcado por una evolución tecnológica dominada por una lógica masculina como un proceso acelerado, sin experimentación previa que caracteriza todo el mundo digital y presenta muchos desafíos éticos, técnicos y sociales.

Mientras el progreso tecnológico ha avanzado de manera exponencial, los cambios en el ámbito social, político y humano son graduales. Tras décadas de transformación digital, la ciudadanía carece de las competencias necesarias que le permitan tomar decisiones informadas y evaluar los riesgos del uso de la inteligencia artificial, así como de los servicios de las plataformas. De igual manera, los poderes públicos y económicos no han afrontado el desafío de regular el sistema, de establecer marcos éticos y legales adecuados y garantizar la soberanía digital. La UNESCO (2020) ya advertía sobre estos problemas y recomendaba integrar la ética y la alfabetización tecnológica en la educación, tanto en el sector público como en el privado y en la sociedad civil, con el objetivo de abordar la desigualdad de género en los principios de la IA.

Mientras, la digitalización acelerada y la improvisación con que se ha implantado ha eludido realizar experimentación y pruebas que permitan evaluar la confiabilidad de los sistemas algorítmicos y de IA. Estos sistemas operan con opacidad, como «cajas negras» que influyen en la selección de contenidos y en la toma de decisiones comerciales sin mecanismos claros de rendición de cuentas. Los riesgos aumentan en materia de seguridad y privacidad, la IA facilita la desinformación y la manipulación informativa a gran escala, sin dejar rastro y sin una regulación adecuada. A esto se suman los efectos sobre la sociedad y la salud mental, pues los algoritmos pueden promover la ansiedad, generar conductas adictivas o intensificar la polarización política.

Como tecnología disruptiva, la IA ha abierto debates sobre la naturaleza de la inteligencia y sus implicaciones éticas y sociales. Aunque no posee un conocimiento real ni conciencia propia, ya que es una simulación basada en modelos matemáticos y procesamiento de datos. Los sistemas de IA pueden parecer inteligentes, pero carecen de comprensión y su capacidad para generar hipótesis y conocimientos se basa en datos históricos, lo que limita su potencial para incorporar innovaciones y cambios disruptivos. Sus aportaciones dependen del contexto en el que opera, y así tiene tendencia a reproducir y amplificar los sesgos preexistentes, especialmente en términos de género, raza y otras categorías interseccionales.

La IA no es neutral ni ajena a las estructuras de poder que rigen la sociedad. Su desarrollo acelerado exige una vigilancia constante para mitigar sesgos y garantizar que sus aplicaciones sean equitativas y éticamente responsables. El lanzamiento de ChatGPT-3 en noviembre de 2022 por OpenAI, inicialmente como una organización sin ánimo de lucro y con

el propósito de generar beneficios para la humanidad, puso en evidencia estas tensiones. La posterior creación de una empresa privada con el mismo nombre para la comercialización de productos de IA ha suscitado debates sobre el equilibrio entre innovación, acceso y control corporativo de estas tecnologías.

3 | El patriarcado y el estándar masculino en la ciencia y la tecnología

La cultura patriarcal en el ámbito tecnológico se ha impuesto con sus valores y prácticas masculinos. Así se han minimizado las aportaciones de las mujeres y de colectivos de la diversidad, consolidando un modelo en el que los hombres ocupan un rol central en el desarrollo, diseño y control de la tecnología (Ensmenger, 2012). Se trata, al igual que en muchos otros ámbitos, del estándar masculino que, tal como señalan teóricas feministas como Simone de Beauvoir y Luce Irigaray, ha definido las experiencias, conocimientos y visiones de los hombres como representativos de la humanidad, excluyendo sistemáticamente a quienes no forman parte de ese colectivo.

Este estándar ha consolidado valores como el control y la racionalidad, entendidos como principios fundamentales del desarrollo tecnológico. Así, los proyectos se conciben como objetivos y analíticos, dejando de lado dimensiones sociales, éticas y emocionales que también son esenciales para responder a las necesidades humanas. Katrine Marçal (2022), en *La madre del ingenio*, explica cómo la narrativa del progreso tecnológico ha estado históricamente vinculada a la masculinidad, equiparada con racionalidad, control y superioridad técnica. La informática, la robótica y la inteligencia artificial no son solo herramientas, sino también símbolos de estatus y poder que refuerzan estructuras patriarcales.

Se ha construido así un relato en el que el progreso aparece como una obra masculina. La inteligencia artificial no escapa a esta lógica: la selección de datos considerados relevantes, las decisiones de diseño y los valores programados en los algoritmos están mediados por una cultura tecnológica que ha priorizado la competitividad y el dominio técnico sobre una perspectiva más inclusiva y humanista del conocimiento. La tecnología se asocia con ideales masculinos como la conquista y el poder, expresados en el dominio sobre el mundo físico, las estructuras sociales y los recursos económicos. Incluso la narrativa sobre el avance tecnológico, al desentenderse de sus impactos en las dimensiones humanas y emocionales, refuerza la idea de que el progreso es una obra de los hombres y, por tanto, un símbolo de virilidad y estatus, especialmente en campos como la informática, la robótica y la inteligencia artificial.

La relación entre los avances tecnológicos y la guerra (Anrubia-Aparici, 2021) refuerza todavía más esta visión masculina. La historia de la tecnología está plagada de metáforas de dominio y violencia, de fuerza y poder militar, hasta el punto de diluir la noción de tecnología como un conjunto de soluciones orientadas a mejorar la calidad de vida. En consecuencia, se han desvalorizado aspectos como el bienestar social, los cuidados, la comunicación o la resolución pacífica de conflictos, campos a los que puede contribuir el potencial transformador de la tecnología.

Lev Manovich, en *Software Takes Command* (2013), analiza cómo el *software* ha devenido el medio dominante en la cultura digital, estructurando nuestras interacciones con la tecnología y la producción de conocimiento. Su concepto de «softwarización» de la sociedad muestra cómo el desarrollo tecnológico no es neutral, sino que está condicionado por estructuras de poder y valores sociales. Esta perspectiva se vincula con los estudios de Cánepa & Ulfe (2014) y Yansen (2020), que señalan cómo el capitalismo y el patriarcado han moldeado la evolución digital desde sus inicios, promoviendo una cultura basada en la eficiencia, la productividad y la competitividad. Manovich sostiene que el *software* no solo media nuestras interacciones, sino que también define las oportunidades y las barreras dentro del sector tecnológico. En este sentido, el paso de la informática de un campo en el que las mujeres desempeñaban roles clave a uno dominado por hombres puede entenderse como parte de un proceso de institucionalización del *software* bajo lógicas de control y exclusión.

Además, Manovich destaca que las herramientas digitales han sido diseñadas para reforzar modelos preexistentes de conocimiento, priorizando enfoques cuantificables y algoritmizados sobre aquellos más creativos o colaborativos. Esto coincide con la crítica de Yansen (2020) sobre la exclusión de las mujeres en espacios de innovación y liderazgo tecnológico, donde las habilidades científicas y matemáticas son privilegiadas mientras se desvalorizan otras formas de conocimiento. El *software* no solo refleja, sino que amplifica, las desigualdades de género en la industria tecnológica, consolidando un modelo que desincentiva la diversidad de enfoques y refuerza la exclusión estructural.

Uno de los ejemplos más evidentes de este sesgo es la cultura *programmer* (Kumar, 2014), que representa la hipermasculinización del mundo digital. En esta narrativa, los hombres aparecen como héroes de la tecnología, promoviendo la individualidad y la competitividad masculina como una cultura que fomenta la idea de una hermandad masculina que premia el éxito basado en largas jornadas laborales y un entorno altamente competitivo, lo que excluye a muchas mujeres que no encajan en ese modelo. Además, asocia el éxito con la combatividad, la homogeneidad masculina y el desprecio por la diversidad de género, lo que se

refleja en espacios como el consumo de videojuegos, la ciencia ficción y el propio mundo del *hacking*.

Son valores ligados a los mitos de la industria tecnológica y a sus creadores como héroes: Bill Gates, Steve Jobs o Elon Musk. Estos personajes han sido retratados como genios creativos con una combinación de habilidades técnicas y liderazgo, reforzando atributos individualistas y capitalistas que contrastan con valores como la colaboración, la experimentación y el trabajo en equipo, tradicionalmente más asociados con lo femenino. En este proceso, se ha invisibilizado el papel fundamental que desempeñaron las mujeres en las primeras décadas del desarrollo de la informática, especialmente en Estados Unidos.

4 | La cultura *brogrammer* y las barreras de género

El patriarcado ha marcado una barrera de género en la educación y el acceso a la digitalización, lo que se traduce en una representación femenina ínfima en la industria tecnológica (Martínez-Cantos & Castaño, 2017). Esta brecha se refuerza a través de la discriminación en la contratación, la falta de credibilidad sobre la capacidad y los logros de las mujeres, y los obstáculos para acceder a redes profesionales dominadas por hombres (Bertogg *et al.*, 2020). Como resultado, la presencia femenina en el sector es menor y se entiende que sus contribuciones tienen menos valor que las masculinas. Esto condiciona el liderazgo, la consolidación de sus propuestas e innovación y refuerza la masculinización de la tecnología.

Esta desigualdad estructural no solo afecta la representación de las mujeres en la industria, sino que la tecnología se construye sin atender adecuadamente sus necesidades, ignorando aspectos como el cuidado y reforzando sesgos de género implícitos en los productos digitales. Los efectos de esta exclusión tienen consecuencias económicas, sociales y políticas, con un modelo caracterizado por su lógica patriarcal y extractiva. La digitalización y el crecimiento de la industria tecnológica han intensificado la concentración de poder en corporaciones y potencias del sector, que operan bajo la lógica de la competitividad, favorece la acumulación de recursos en manos de unos pocos, y consolida estructuras de exclusión y desigualdad.

El individualismo, la competitividad extrema y la maximización del rendimiento personal excluyen a quienes no se ajustan a los roles tradicionales del sector, en particular a las mujeres y a las minorías. En el ámbito laboral, esta dinámica da lugar a entornos tóxicos y hostiles, caracterizados por actitudes machistas, comentarios sexistas y conductas discriminatorias (Levin, 2017). El uso de un lenguaje despectivo hacia las mujeres y la exaltación de una masculinidad agresiva refuerzan esta exclusión.

La presión constante por el éxito económico, la rapidez en el desarrollo y la búsqueda incesante de la novedad (más que de la innovación) desplazan valores fundamentales en el trabajo, como el bienestar, la inclusión y la diversidad. La mercantilización del sector, centrada en la productividad y el beneficio, refuerza la hegemonía masculina y perpetúa la marginación de las mujeres. Este proceso está acompañado de un elitismo tecnológico creciente, en el que el concepto de eficiencia se ajusta exclusivamente al perfil de la cultura *brogrammer*, excluyendo cualquier perspectiva que incorpore valores asociados a la diversidad o al enfoque colaborativo.

La exclusión de las mujeres no solo se refleja en la estructura laboral de la industria digital, sino también en el desarrollo de productos tecnológicos. El mercado, la publicidad y la industria del deseo han impulsado un diseño de productos centrado en las experiencias y preferencias masculinas, en lugar de priorizar los problemas sociales o las necesidades de colectivos vulnerables (García-García, Felip-Miralles & Galán-Serrano, 2017). Cuando las mujeres no están presentes en los procesos de diseño, concepción de productos o programación, sus necesidades e ideas no son adecuadamente consideradas, lo que genera soluciones tecnológicas que no satisfacen a una parte significativa de la población. Además, la dinámica del mercado refuerza esta desigualdad al asumir que los principales consumidores de los productos tecnológicos son hombres y, por tanto, diseñar específicamente para ellos.

La segmentación del mercado laboral por género ha sido particularmente evidente en sectores de alta rentabilidad y prestigio social, como la industria digital en las últimas décadas (Rodríguez-Escanciano, 2024). Esta segregación limita el acceso de las mujeres a posiciones de liderazgo, mejores retribuciones y participación en la innovación y el desarrollo tecnológico. Al valorarse las contribuciones en función de su impacto económico, las aportaciones masculinas no solo adquieren mayor reconocimiento financiero, sino también un mayor prestigio dentro del sector.

5 | La violencia simbólica y estructural del sexismo digital

En todos los fenómenos y procesos descritos, especialmente en la cosificación de las mujeres, subyace un sistema de violencia simbólica y estructural que sintetiza el sexismo digital. Este sexismo se manifiesta de manera evidente, aunque a menudo pase desapercibido, en las industrias digitales y en las prácticas comunicativas actuales.

En primer lugar, la brecha digital de género: la desigualdad en el acceso y el uso de la tecnología es un círculo vicioso, ya que la falta de mujeres en los procesos creativos e innovadores del diseño de productos provoca que no se contemplen las necesidades y realidades de las

mujeres. Como ejemplo, se puede citar el diseño de aplicaciones de salud que no contemplan variables específicas de las mujeres (Moulaei *et al.*, 2023). Esto provoca que se perpetúen las desigualdades de uso y que se mantenga en el acceso, pero también en los beneficios que aportan los productos digitales. La brecha se abre en la cultura, se mantiene en el sistema educativo y en la industria digital, se refuerza por la falta de referentes y modelos, por la discriminación en el trabajo y por la hostilidad del modelo hacia las mujeres.

En segundo lugar, la representación y los estereotipos de género: el sesgo masculino en el diseño tecnológico conlleva la infrarrepresentación de otras identidades de género (García-García, Felip-Miralles & Galán-Serrano, 2017) y limita la percepción de sus capacidades y roles. Además, se refuerza un estándar inalcanzable de la belleza o del comportamiento que limita el empoderamiento y las oportunidades de las mujeres al aceptar modelos irreales en redes sociales, publicidad o videojuegos, que imponen narrativas perjudiciales. De hecho, la cosificación de las mujeres en el mundo de las representaciones sociales y comunicativas como en la propia tecnología se puede entender tanto como un mecanismo de poder, a través de la violencia simbólica para mantener este estado de la cuestión o como una consecuencia del mismo proceso. Tanto los asistentes virtuales (Torres-Guijarro, 2023) como los videojuegos (Curiel, 2018) reflejan el imaginario masculino en el que las mujeres son subordinadas, accesorias o bien objetos de placer y disfrute, usuarias pasivas, figuras secundarias, pero nunca sujetos de derecho y de igualdad. El lenguaje y las microagresiones también pasan desapercibidas cuando, sobre todo en espacios dominados por hombres, como el mundo *gamer*, se proponen roles tradicionales para las mujeres o representaciones sexualizadas (Tabares, 2023).

El desarrollo de la inteligencia artificial ha venido a consolidar las desigualdades de género y etnia en el reconocimiento y en la generación de imágenes, ya que tanto la reproducción de los modelos de visión artificial y de sus resultados, como la influencia de quienes participan en la clasificación de imágenes contribuyen a ampliar el sesgo existente (Kafkalias *et al.*, 2022).

En la industria cinematográfica y audiovisual, se han desarrollado sistemas de detección de los sesgos en la representación en las películas y existen numerosos tests que ayudan a determinar el peso de estos sesgos, como el test de Bechdel. Se suman estudios más detallados que establecen porcentajes sobre las diferencias de género en los protagonistas de las películas (Davis, 2019). En el estudio que se realizó sobre cien películas taquilleras entre 2007 y 2017, los datos son muy representativos: cuando los hombres son los protagonistas indiscutibles de las películas su tiempo en pantalla es de un 77% frente a un 23% de las mujeres. Y su tiempo de diálogo o *speaking time* es de un 73% frente a un 27%. Si ellas eran las protagonistas, su tiempo de pantalla respecto a la de ellos era

de un 52% frente a un 48%. Es decir, aparecen prácticamente el mismo tiempo. Es interesante analizar cómo estos sesgos del cine se reproducen en el mundo digital. Un estudio sobre Twitter reveló que las narrativas de los tuits se asemejan más a las películas que no superan el test de Bechdel que a las que sí lo hacen (García *et al.*, 2014).

En tercer lugar, las plataformas digitales, en su búsqueda de maximizar beneficios, refuerzan sesgos preexistentes. Al estar diseñadas principalmente por hombres, perpetúan estereotipos que afectan desde la recomendación de contenidos hasta el desarrollo de la IA en diversos ámbitos (O'Neil, 2018). De esta forma, se perpetúan los roles tradicionales en los contenidos recomendados como noticias o productos culturales, que, a su vez, reducen la presencia de mujeres creadoras, ya que los algoritmos potencian la carrera de hombres a costa de las oportunidades para las mujeres.

Los algoritmos potencian carreras masculinas a costa de las oportunidades de las mujeres y perpetúan roles tradicionales. Plataformas de contratación automatizada favorecen a los hombres porque sus modelos de entrenamiento reflejan un historial de contratación sesgado (Bravo-Villasante, 2023). Un caso emblemático ocurrió en 2018, cuando el sistema de IA de reclutamiento de Amazon discriminaba sistemáticamente a las mujeres. Entrenado con datos históricos, penalizaba currículums con términos asociados a mujeres y favorecía aquellos con vocabulario más comúnmente masculino (Kodiyani, 2019; Dastin, 2022). De la misma forma, en España, experimentos en la plataforma Infojobs evidenciaron sesgos en la selección de profesionales de la comunicación (Delgado Bancalero, 2023).

El sesgo algorítmico se refuerza en las prácticas y en el uso de plataformas digitales; afecta a la mayoría de los servicios digitales, desde búsquedas a redes sociales, y al diseño y desarrollo del *software*, a la inteligencia artificial, así como a los principales modelos de aprendizaje autónomo que los amplifican. También alcanza a los sistemas automatizados de toma de decisiones dentro de las administraciones, ya que cada vez se realizan más procesos de selección de esta forma, tanto para el acceso a servicios sociales, como para concesiones de créditos, de becas o para la aplicación de beneficios como los de los seguros médicos o para procedimientos administrativos (Gamero, 2023). La mayoría de los algoritmos que se utilizan no son transparentes ni auditables, resulta imposible conocer los criterios de la selección y tienen sesgos importantes relacionados con el género y la interseccionalidad. El Observatorio de Algoritmos con Impacto Social (OASI) de la Fundación Ética (2023) ha diseñado una herramienta para auditar algoritmos y determinar sus sesgos¹, que

1. <https://airtable.com/appkg5eQBvy3zcosd/shrsAN2oTf68kM6O9/tblG2604tSoMOcwWX?backgroundColor=teal&viewControls=on>

se encuentra disponible, creado un registro con información sobre determinados algoritmos (nombre, descripción, localización, objetivo, impacto social, etc.) para aumentar la concienciación pública (Godino *et al.*, 2022)

En cuarto lugar, el diseño de productos tecnológicos también contribuye a la exclusión de las mujeres. Estudios han demostrado que interfaces, dispositivos y aplicaciones suelen estar diseñados para patrones de uso masculinos. Esta norma implícita prioriza intereses de los hombres e invisibiliza la experiencia femenina.

Se presentan barreras que limitan su presencia e influencia en la creación de tecnologías y en el acceso al trabajo, y a su independencia económica, no solo por el beneficio directo de la participación en plataformas digitales, sino también a través de los algoritmos o la inteligencia artificial que favorecen los contenidos diseñados por hombres, así como su mayor visibilidad en redes sociales. Se conoce como «techo de cristal algorítmico» (Stoica *et al.*, 2018) a la barrera social metafórica que impide que mujeres o personas de color alcancen una representación igualitaria en Instagram. El efecto se ha investigado a través de un modelo matemático que muestra cómo los algoritmos de recomendación social pueden aumentar la infrarrepresentación de ciertos grupos demográficos incluso en la parte superior de la jerarquía social, donde las mujeres obtienen una representación del 25% frente al 75% masculino. En las prácticas en redes sociales hay que añadir la *manosfera* como un ámbito hostil en el que se reproducen lógicas sociales de hostigamiento como el *mansplaining* o el silenciamiento, la interrupción permanente a mujeres que las disuade de la participación y las invisibiliza. Este contexto también limita la participación de las mujeres por el efecto de las campañas de acoso, de desinformación y el temor a las represalias (Arecón-Beltrán *et al.*, 2022; Piñero Otero *et al.*, 2024). E incluso se agrava por el uso de los memes de la *manosfera*, que hacen trivial el acoso o la discriminación de género y buscan en el humor un refuerzo para las ideas sexistas (Drakett, 2018).

Este fenómeno tiene un impacto significativo, ya que la norma masculina domina el desarrollo de IA y tecnologías emergentes. Asistentes virtuales, por ejemplo, perpetúan la idea de que las mujeres están para servir, reforzando estereotipos patriarcales. Además, las voces de estos asistentes no solo replican prejuicios, sino que también sufren acoso verbal, lo que refleja una violencia simbólica en la interacción digital (Cernadas & Calvo-Iglesias, 2020).

En quinto lugar, el diseño de plataformas digitales no suele contemplar necesidades específicas de privacidad y seguridad para las mujeres, quienes son más vulnerables al acoso, la explotación y el fraude. Son víctimas de formas específicas de violencia como el *doxing* (publicación de información privada), el *revenge porn* (difusión no consentida de contenido sexual) y el ciberacoso (Barros de Santana *et al.*, 2024).

Muchas *apps* de seguimiento menstrual han sido criticadas por priorizar la fertilidad y la planificación familiar sobre la salud general de las mujeres y olvidar la salud reproductiva. Pero también por la falta de privacidad y la venta de datos sin consentimiento (Epstein *et al.*, 2017), por la falta de investigación y profesionalidad en su desarrollo (Earle *et al.*, 2021). Desde que en 2022 se perdiera la protección constitucional en el derecho al aborto en Estados Unidos, que desembocó en que numerosos estados hayan prohibido la interrupción del embarazo, resultan vulnerables las mujeres cuyos datos comparten este tipo de aplicaciones. El uso de estas aplicaciones se ha convertido en un riesgo, ya que recopilan datos altamente sensibles que pueden ser utilizados en contextos de persecución legal (Rightley, 2024).

Conclusiones

El sexismo digital es un fenómeno estructural que no se limita a las manifestaciones explícitas en plataformas digitales, sino que se encuentra arraigado en el desarrollo mismo de la tecnología y de la inteligencia artificial. La cultura patriarcal en el ámbito tecnológico ha consolidado un modelo basado en valores y prácticas masculinas que refuerzan la desigualdad de género, marginando las contribuciones de mujeres y colectivos diversos, por lo que las tecnologías no solo reproducen, sino que amplifican los sesgos de género y producen una serie de consecuencias en el acceso, la representación y el uso de los productos digitales.

El desarrollo acelerado de la IA sin la debida experimentación ha dado lugar a sistemas opacos que operan sin rendición de cuentas, favoreciendo la exclusión de perspectivas inclusivas y humanistas. La falta de marcos éticos y regulatorios adecuados ha permitido que los sistemas de IA perpetúen desigualdades, favoreciendo modelos que no contemplan la diversidad y la equidad. En este contexto, las recomendaciones de organismos como la UNESCO, que enfatizan la necesidad de integrar la ética y la alfabetización tecnológica y algorítmica en la educación, resultan fundamentales para abordar estos problemas desde una perspectiva estructural.

Asimismo, la consolidación de una cultura hipercompetitiva y masculinizada en el sector tecnológico ha contribuido a la exclusión de mujeres en la educación, la industria digital y los espacios de liderazgo. Este fenómeno, la cultura *brogrammer*, refuerza la discriminación y dificulta la participación de mujeres en procesos de innovación y toma de decisiones. Como resultado, los productos digitales y las aplicaciones tecnológicas se diseñan sin atender adecuadamente las necesidades de las mujeres, perpetuando desigualdades estructurales y consolidando un mercado que privilegia las experiencias y expectativas masculinas.

La violencia simbólica y estructural del sexismo digital también se manifiesta en la brecha de género en el acceso y uso de la tecnología, en la representación de las mujeres en la cultura digital y en los estereotipos de género en los productos y servicios tecnológicos.

Ante este panorama, es imperativo desarrollar estrategias que promuevan una tecnología más inclusiva, regulaciones que mitiguen los sesgos algorítmicos y políticas que fomenten la participación equitativa de mujeres y colectivos diversos en el sector tecnológico. Solo mediante una transformación profunda de las estructuras que sostienen el sexismo digital será posible avanzar hacia una sociedad donde la tecnología se convierta en una herramienta de equidad y justicia social. Por ello, se hace imprescindible una mayor transparencia en el diseño de algoritmos, auditorías en plataformas digitales y una perspectiva de género en la innovación tecnológica.

Todas estas cuestiones tienen un enorme impacto en la autonomía y empoderamiento de las mujeres, ya que son cuestiones que afectan a lo cotidiano, se disipan de forma sutil en actitudes naturalizadas y se normalizan a partir de las prácticas de comunicación. Así, vemos en las plataformas digitales cómo se consolidan estereotipos de género en las formas de interacción, en la interfaz o en aplicaciones, como, por ejemplo, los filtros de belleza para las fotografías que promueven estándares estéticos tradicionales y la cosificación de las mujeres.

Muchos de los aspectos que caracterizan el sexismo digital forman parte del patriarcado y de la cultura capitalista que lo reproduce, como la competitividad, el ejercicio del poder y una lógica de la desigualdad que mantiene las barreras entre lo público y lo privado, en especial en la mercantilización del trabajo. Luchar contra estas dinámicas resulta difícil por cuánto implica la necesidad de una intervención multidisciplinar a través de la educación, la política y la legislación contra las prácticas hostiles y la implicación del sector tecnológico en la creación de contextos más inclusivos y seguros para las personas, tanto en el mundo laboral, como en los escenarios de la comunicación online.

Referencias

- Arecón-Beltrán, S., Sola-Morales, S. y Hernández-Conde, M. (2022). Fascismo digital para bloquear la participación y la deliberación feminista. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 20(1), 25-35. <https://doi.org/10.5209/tekn.81002>.
- Ávila Bravo-Villasante, M. (2023). La agenda feminista ante la cuarta revolución industrial. Mujeres y algoritmización de la esfera pública. *Cuestiones de Género: de la Igualdad y la Diferencia*, 18, 132-155. <https://bit.ly/40NZGtX>.

- Barros de Santana, M., Neumann Fernandes, R. y da Silva Sousa, J. (2024). EXPOSIÇÃO À INTIMIDADE DA MULHER, SEM AUTORIZAÇÃO, POR PARTE DO EX-COMPANHEIRO: REVENGE PORN. *Revista Ibero-Americana de Humanidades, Ciências e Educação- REASE*, 10(5). <https://doi.org/10.51891/rease.v10i5.13972>.
- Bertogg, A., Imdorf, C., Hyggen, C., Parsanoglou, D. y Stoilova, R. (2020). Gender Discrimination in the Hiring of Skilled Professionals in Two Male-Dominated Occupational Fields: A Factorial Survey Experiment with Real-World Vacancies and Recruiters in Four European Countries. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 72, 261-289. <http://bit.ly/3WvvgIKU>.
- Cânepa, G. y Ulfe, M. (2014). Performatividades contemporáneas y el imperativo de la participación en las tecnologías digitales. *ANTHROPOLOGICA*, 22(33), 67-86. <https://doi.org/10.18800/anthropologica.201402.004>.
- Cernadas, E. y Calvo-Iglesias, E. (2020). Gender perspective in artificial intelligence (AI). *Eighth international conference on technological ecosystems for enhancing multiculturalism*, 173-176. <https://doi.org/10.1145/3434780.3436658>.
- Curiel, Á. (2018). La mujer como actante. En Amores, M. (coord.): *¡Protesto! Videojuegos desde una perspectiva de género* (111-137). Anait games.
- Drakett, J. et al. (2018). Old Jokes, New Media—Online Sexism and Constructions of Gender in Internet Memes. *Feminism and Psychology*, 28(1), 109-127. <https://bit.ly/4gh1vUK>.
- Ensmenger, N. (2012). *The Computer Boys Take Over*. The MIT Press.
- García-García, C., Felip-Miralles, F. y Galán-Serrano, J. (2017). Diseño, producto y usuario: el papel del consumidor en el proceso de co-creación. *Revista Sonda: Investigación en Artes y Letras*, (6), 123-130. <https://doi.org/10.4995/sonda.2017.18382>.
- Levin, S. (4 de marzo de 2017). Silicon Valley, el reino de la tecnología lo es también del machismo y la discriminación. *eldiario.es*. https://www.eldiario.es/internacional/theguardian/silicon-valley-centro-acoso-sexual_1_3552405.html.
- Martínez-Cantos, J. L. y Castaño, C. (2017). La brecha digital de género y la escasez de mujeres en las profesiones TIC. *Panorama social*(25), 49-65. https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_PS/025art05.pdf.
- Moulaei, K., Moulaei, R. y Bahaadinbeigy, K. (2023). Barriers and facilitators of using health information technologies by women: a scoping review. *BMC Medical Informatics and Decision Making*. Doi: <https://doi.org/10.1186/s12911-023-02280-7>.
- O'Neil, C. (2018). *Armas de destrucción matemática: Cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Capitán Swing.
- Piñero Otero, T., Martínez Rolán, X. y Castro Souto, L. M. (2024). ¿Sueñan los troles con mujeres en el poder? Una aproximación al troleo de género como violencia política. *Anuario Electrónico de Estudios en Comunicación Social «Disertaciones»*, 17(2). <https://doi.org/10.12804/revistas.urosario.edu.co/disertaciones/a.13988>.
- Rodríguez-Escanciano, S. (2024). La brecha digital de género: reflexiones en las pos-trimerías del primer cuarto del siglo XXI. *Revista de Trabajo y Seguridad Social. CEF*(479), 7-19. <https://doi.org/10.51302/rtss.2024.21415>
- Stoica, A. A., Riederer, C. y Chaintreau, A. (2018). Algorithmic Glass Ceiling in Social Networks. *Track: Web and Society*, 23-27. <https://doi.org/10.1145/3178876.3186140>.

- Tabares, M. S. (2023). The Hypersexuality Of Women and Videogames: The Effects It Has on Society and The Business Industry. *UC Merced Undergraduate Research Journal*. <https://doi.org/10.5070/M416162640>.
- Torres-Guijarro, S. (2023). Sesgos de género de la asistencia digital en español. *Gender on Digital. Journal of Digital Feminism*, 1, 35-58. <https://doi.org/10.35869/god.v1i.5061>.
- Yansen, G. (2020). Género y tecnologías digitales: ¿qué factores alejan a las mujeres de la programación y los servicios informáticos? *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 17(2), 239-249. <https://doi.org/10.5209/tekn.69472>.

Ética y transparencia: claves del uso responsable de la inteligencia artificial en Radio y Televisión Española (RTVE)

Yasir Barceló

Universidad de Málaga

yasir@uma.es

Nereida Cea

Universidad de Málaga

nereidacea@uma.es

Sonia Blanco

Universidad de Málaga

sblanco@uma.es

1 | Introducción

La inteligencia artificial (IA) ha experimentado un significativo avance desde sus primeras formulaciones teóricas hasta llegar a ser una herramienta fundamental en diversos ámbitos, incluidos los medios de comunicación. Su desarrollo no solo ha transformado la forma en que se produce y consume información, sino que también ha conllevado el planteamiento de preguntas fundamentales sobre su regulación ética y social, ya sea en su uso productivo o como herramienta de detección de noticias falsas. En el contexto de Radio y Televisión Española (RTVE), un medio de

titularidad pública con un inequívoco compromiso con la transparencia y la responsabilidad social, estas cuestiones son especialmente relevantes (Mayoral Sánchez *et al.*, 2023).

La concepción moderna de la IA parte de los trabajos iniciales de Alan Turing y su famoso *Computing Machinery and Intelligence* (1950), en el que planteó la posibilidad de que las máquinas pudieran «pensar». Además de formular el conocido *Test de Turing* como un criterio operativo para evaluar la inteligencia máquina, también sentó las bases para un campo de estudio que en sus inicios se centraba en la resolución de problemas lógicos y matemáticos. Durante la conferencia de Dartmouth en 1956, considerada el punto de partida oficial de la IA, se acuñó el término «inteligencia artificial» y se abrieron caminos destinados a la investigación multidisciplinar, involucrando a matemáticos, psicólogos y científicos de la computación (Russell y Norvig, 1995).

Los primeros avances en IA se centraron en resolver problemas cerrados y bien definidos, como el desarrollo de programas que podían jugar al ajedrez o demostrar teoremas matemáticos. Sin embargo, en los años 70 se llegó a lo que se conoció como el «invierno de la IA», por el estancamiento producido ante la falta de avances tecnológicos y teóricos que permitiesen nuevos logros, debido a la pérdida de interés y fuentes de financiación. Aun así, las décadas siguientes presenciaron, gracias al desarrollo de sistemas expertos en los años 80, como *Mycin* y *Dendral* (McCorduck, 1979), que demostraron el potencial de la IA aplicada a dominios específicos como la medicina.

Con el avance del *hardware* y las redes neuronales en las décadas de 1990 y 2000 comenzó a expandirse, llegando a usos comerciales y cotidianos, más allá de las aplicaciones meramente científicas. Tecnologías como el reconocimiento de voz, la visión a través de ordenadores y los sistemas de recomendación transformaron sectores como la industria, el entretenimiento y las telecomunicaciones (Parratt-Fernández *et al.*, 2024). De forma paralela, esto conlleva el surgimiento de debates y preocupaciones éticas en relación con el impacto de estas nuevas tecnologías en la privacidad, la estabilidad del empleo de los profesionales y la idoneidad de la toma automatizada de decisiones.

La década de 2020 trajo consigo la explosión definitiva de la IA generativa, lo que constituye un nuevo e importante hito en el desarrollo de esta tecnología. Se va más allá del almacenamiento y ordenamiento de datos, llegando a la creación de contenidos originales, con herramientas como ChatGPT o DALL-E, lo que abre un abanico de posibilidades cada vez más amplio (Gómez-Calderón y Ceballos, 2024). Todo ello aumenta el debate sobre los riesgos de uso de esta tecnología emergente, ya que la capacidad de estas tecnologías para generar textos, imágenes y vídeos indistinguibles de los creados por humanos ha planteado nuevos desafíos en áreas como la desinformación y la manipulación mediática.

En términos de regulación, la Unión Europea ha liderado los esfuerzos para establecer marcos legales que guíen su desarrollo y uso. La propuesta de Ley de IA, presentada en 2021, clasifica las aplicaciones de inteligencia artificial según su nivel de riesgo y establece requisitos muy estrictos para sistemas de alto impacto, como los usados en reconocimiento facial o sistemas de toma automatizada de decisiones (Valencia y Túñez, 2018). De esta manera se busca un equilibrio entre la innovación y la protección de derechos fundamentales, siendo este un enfoque que es secundado por RTVE en su estrategia de integración y uso de la IA.

Sin embargo, la falta de una regulación homogénea a nivel global provoca que se produzcan desigualdades entre las grandes potencias mundiales a la hora de afrontar los retos que se plantean. Mientras que Estados Unidos y China avanzan con enfoques más flexibles, Europa enfrenta el reto de mantenerse competitiva y seguir a la vanguardia de los nuevos alcances de la tecnología sin comprometer sus principios éticos. Esta tensión se refleja en el debate interno de RTVE, donde se aviva el dilema entre el progreso y una mayor autorregulación a través de códigos éticos internos que complementen las normativas existentes.

En el ámbito de los medios de comunicación en particular, la IA se ha convertido en una herramienta útil para mejorar la eficiencia y personalización de los contenidos, así como para automatizar tareas tediosas y permitir a los profesionales centrarse en otros aspectos de su trabajo. Sin embargo, también ha generado preocupaciones éticas y sociales, especialmente en cuanto a su impacto en el trabajo periodístico, la estabilidad de los puestos de trabajo y la veracidad de la información (Guallar *et al.*, 2020). En este punto, y como medio público, Radio y Televisión Española (RTVE) tiene la responsabilidad de incorporar estas tecnologías de manera ética, con el objetivo de garantizar la transparencia en su implementación y fomentar la confianza de la audiencia. La corporación ha apostado por integrar la IA en procesos clave como la generación automática de noticias, la verificación de información y la automatización de tareas, como el subtítulo de programas. Estas iniciativas buscan mejorar la eficiencia, adaptarse a las posibilidades que ofrece el desarrollo tecnológico y fortalecer el compromiso de RTVE con la inclusión y la calidad informativa.

En definitiva, la evolución de la IA desde sus orígenes hasta su estado actual ilustra una trayectoria marcada por la búsqueda del equilibrio entre los avances tecnológicos y los dilemas éticos que suscitan su implementación y uso. En el caso de RTVE, este hecho se realza aún más dada su posición de servicio público, que exige un enfoque responsable y transparente en la adopción de tecnologías emergentes (Real *et al.*, 2024). El desarrollo continuo de marcos deontológicos éticos y la promoción de la alfabetización digital serán esenciales para garantizar que la IA se convierta en una herramienta para fortalecer la confianza y la inclusión en los medios de comunicación, frente a los usos inadecuados de los que también se hace uso.

2 | Revisión de la literatura: marco teórico

La incorporación de la IA en los medios de comunicación plantea importantes retos en el ámbito ético. Este proceso requiere un enfoque que garantice el respeto a principios fundamentales como la privacidad y la transparencia. De acuerdo con González y Sanahuja (2023), la irrupción de la IA en el periodismo ha transformado los métodos de trabajo, comprometiendo en ocasiones la calidad informativa y la excelencia periodística. Además, se han identificado riesgos éticos que exigen una autorregulación responsable para evitar sesgos algorítmicos y asegurar prácticas adecuadas en la producción de contenido (González y Sanahuja, 2023).

En el ámbito de la comunicación, los principios de tecnoética han adquirido relevancia. Esta disciplina emergente analiza las interacciones entre tecnología, ética y sociedad, promoviendo el uso responsable de sistemas como la IA para garantizar la justicia y la calidad informativa (Aguilera, 2023).

La IA ha comenzado a desempeñar un papel crucial en el periodismo moderno, facilitando la automatización de tareas como la generación de noticias, el análisis de grandes volúmenes de datos y la verificación de información. Sin embargo, también ha generado debate y controversia en lo referido a la desinformación y el sesgo algorítmico. En este sentido, Quimis (2024) destaca que la colaboración entre humanos y sistemas de IA es clave para mejorar la precisión y la velocidad en la difusión de noticias, mientras que se evita la amplificación de prejuicios existentes. Además, Quian y Sixto (2024) encontraron que, aunque la IA generativa aún no se aplica de forma masiva en los medios, existe un interés creciente por su implementación, destacándose la falta de expertos en esta tecnología como un desafío clave. Según Palomo *et al.* (2023), la IA no solo se percibe como una herramienta para agilizar procesos productivos en el periodismo, sino también como un recurso clave para combatir la desinformación. Sin embargo, su integración plantea dilemas éticos, especialmente en la generación automatizada de contenido y la alteración de la realidad, lo que exige una reflexión profunda sobre su uso responsable y adecuado.

Por otro lado, el uso de tecnologías como ChatGPT y DALL-E (y todos los modelos surgidos con posterioridad) en la generación de contenido ha transformado la manera en que las noticias son producidas y consumidas. Sin embargo, Vargues-Vara (2024) señala que estas innovaciones también han aumentado los riesgos de desinformación y la pérdida de credibilidad en los medios, subrayando la necesidad de renovación de principios deontológicos.

En Europa, la Ley de IA de la Unión Europea, aprobada en 2024, es un nuevo referente en la regulación del uso de esta tecnología. Esta

legislación, cuyo debate comenzó en 2021, clasifica las aplicaciones de IA según su nivel de riesgo y fija requisitos estrictos para los sistemas de alto impacto. Según Villalobos (2023), esta normativa incluye medidas de gobernanza y supervisión para garantizar la ética y la transparencia en su uso y regula sistemas de alto riesgo mediante auditorías y principios de trazabilidad. Además, la versión final de la ley incluye nuevas reglas en áreas como la vigilancia biométrica y la toma de decisiones automatizadas (Parlamento Europeo, 2024).

Estas normativas buscan equilibrar la protección de derechos fundamentales con la innovación tecnológica, un enfoque que RTVE ha adoptado en sus estrategias de integración de la IA (Lopezosa *et al.*, 2023). En este contexto continental, Montiel Torres *et al.* (2024) subrayan que los países del sur de Europa, como España y Grecia, enfrentan niveles bajos de confianza en los medios, destacando la necesidad de marcos regulatorios que fortalezcan la credibilidad pública, un reto también presente en otros contextos globales como China y Estados Unidos.

Mientras tanto, en América Latina, los esfuerzos reguladores han sido más dispersos. Aguilera (2023) subraya que, aunque la IA ha comenzado a ser adoptada en el periodismo latinoamericano, persisten brechas significativas en su implementación ética debido a la falta de marcos regulatorios uniformes.

En este contexto, los medios públicos afrontan una mayor responsabilidad: adoptar tecnologías emergentes y garantizar un uso ético y responsable de las mismas. Este es el caso de RTVE, que, como se ha mencionado con anterioridad, utiliza IA en la generación automática de noticias, la verificación de información y la automatización de tareas como el subtítulo de programas. Estas iniciativas han permitido mejorar la eficiencia y promover la inclusión, en línea con los valores de servicio público (Vargues-Vara, 2024). Y, concretamente, en sus programas informativos han evolucionado hacia un enfoque pedagógico frente a la desinformación, reforzando su misión de servicio público y su compromiso ético en la lucha contra los desórdenes informativos (Blanco *et al.*, 2023).

Además, RTVE también ha adoptado códigos deontológicos internos para complementar las normativas europeas, fomentando una cultura de transparencia y confianza. Investigadores como González-Esteban sugieren que estas prácticas podrían servir de modelo para otros medios de comunicación públicos (2023).

La integración en los medios de comunicación representa un avance significativo, pero también un desafío ético y regulatorio. En el caso de RTVE, el equilibrio entre innovación y responsabilidad ética se convierte en un ejemplo relevante para la industria. La creación de marcos deontológicos claros, el cumplimiento de normativas y la promoción de la alfabetización digital serán esenciales para garantizar un futuro donde la IA fortalezca la confianza pública y la calidad informativa.

3 | Metodología

Este estudio se desarrolla bajo una metodología cualitativa y se centra en un análisis de caso sobre la implementación, desarrollo y situación actual del uso de la IA en Radio Televisión Española (RTVE). Para ello, se efectuaron dos entrevistas semiestructuradas: una con David Corral, responsable del Departamento de Innovación, y otra con Silvia García, coordinadora de los Centros Territoriales. La selección de estos perfiles respondió a la premisa de elegir periodistas dentro del espectro de profesiones involucradas en esta área del ente público.

El objetivo principal es, a través de las experiencias y opiniones de profesionales destacados en RTVE, responder a las siguientes preguntas de investigación (PI):

- PI1: ¿Qué enfoques se están adoptando a nivel internacional para abordar los dilemas éticos y las regulaciones relacionadas con la IA?
- PI2: ¿De qué manera RTVE integra principios éticos y normativos en su uso de IA dentro del ámbito periodístico y organizacional?

La entrevista con David Corral se llevó a cabo el 8 de mayo de 2024 mediante videoconferencia, previa autorización para su grabación y transcripción. En lugar de plantear las preguntas de investigación de forma literal, se sostuvo una conversación de 32 minutos basada en doce preguntas clave, distribuidas en cuatro bloques temáticos: uso actual de la IA en el periodismo, estructura profesional en RTVE para su desarrollo, marco legal y geográfico, y perspectivas de futuro según los profesionales y los medios. En todos los bloques se abordaron aspectos relacionados con la ética y la regulación.

Por su parte, el diálogo con Silvia García tuvo lugar el 29 de mayo de 2024, siguiendo las mismas directrices técnicas y temáticas. En este caso, el contacto se consiguió mediante la técnica de muestreo bola de nieve (Martín-Crespo y Salamanca, 2007), facilitado por el propio David Corral.

La metodología empleada en este trabajo incluye el uso de herramientas de IA como apoyo en la elaboración de las referencias bibliográficas (ChatGPT 4.0) y para la transcripción de las entrevistas (Microsoft Teams).

4 | Discusión y resultados

4.1. Análisis comparativo

El uso de la IA en RTVE refleja diferencias clave en la forma en que los entrevistados, David Corral y Silvia García, abordan la ética y la regulación

en el uso de esta tecnología. Mientras Corral ofrece una visión optimista centrada en la innovación tecnológica y su aplicación responsable, García pone énfasis en la transparencia y la autorregulación para generar confianza en el público.

Corral destaca la IA como una «herramienta que complementa al periodista», permitiéndole no ocuparse de tareas tediosas, mejorando así su productividad. Por el contrario, García subraya la importancia de que el público conozca cómo y cuándo se utiliza esta tecnología, especialmente en el caso de herramientas generativas. Ambos coinciden en la percepción de que la IA no debe reemplazar a los profesionales, sino fortalecer el periodismo de calidad partiendo de un uso adecuado de esta tecnología.

En cuanto a la regulación, Corral resalta la necesidad de cumplir con las normativas europeas, como la Ley de IA de la Unión Europea, y asegura que RTVE ya trabaja con altos estándares en ética y control de contenidos. Por su parte, García considera esencial un código deontológico interno que complemente las legislaciones existentes, dotando a RTVE de una guía clara y concisa para el uso responsable de la IA.

Un aspecto común es la lucha contra la desinformación. Corral y García coinciden en que puede ser un instrumento de verificación de información y generación de contenidos fiables. Sin embargo, ambos reconocen que combatir la desinformación requiere no solo tecnología, sino también educación del público, no solo en el uso de las herramientas sino también en la percepción y lectura de la información, así como una gestión ética de los datos.

Finalmente, en cuanto a la estructura organizativa, Corral enfatiza la colaboración entre diferentes perfiles profesionales dentro de RTVE, mientras que García destaca la necesidad de coordinar esfuerzos entre los centros territoriales y el Departamento de Innovación. Este enfoque diferente se debe al puesto que desempeña cada uno en el organigrama de RTVE, lo cual les hace colaborar y coordinarse de manera organizada. Ambos consideran que la diversidad de perfiles y la colaboración multidisciplinaria son fundamentales para el éxito de los proyectos de IA en la corporación.

4.2. Respuesta a las preguntas de investigación

PI1: ¿Qué enfoques se están adoptando a nivel internacional para abordar los dilemas éticos y las regulaciones relacionadas con la IA?

Los entrevistados señalan que la regulación internacional, aunque puede considerarse que se encuentra en un estado inicial, ya ha generado acuerdos significativos como la Declaración de Bletchley (2023). Corral

considera que estos marcos son imprescindibles para asegurar un uso responsable de la IA, especialmente frente a las amenazas de desinformación y el mal uso de los datos personales. Sin embargo, también advierte que el exceso de regulación podría ralentizar la innovación en Europa, especialmente frente a China y Estados Unidos.

Por su parte, García destaca la falta de una legislación global homogénea como un desafío clave, lo que refuerza la necesidad de autorregulación. En su opinión, RTVE debe liderar la implementación de códigos deontológicos internos que sirvan como referencia para otros medios de comunicación, tanto públicos (a niveles autonómicos y europeos) como privados.

PI2: ¿De qué manera RTVE integra principios éticos y normativos en su uso de IA dentro del ámbito periodístico y organizacional?

RTVE adopta un enfoque proactivo para garantizar la ética en el uso de la IA. Corral menciona que la corporación ya cumple con las normativas europeas y opera con altos estándares en transparencia y control de contenidos. En el ámbito operativo, se utiliza para tareas como la generación automática de noticias y la verificación de datos, siempre, de momento, con supervisión humana.

García complementa esta visión al enfatizar la importancia de que RTVE desarrolle su propio marco ético. Considera que la transparencia es clave para mantener la confianza del público y sugiere que cualquier contenido generado con IA debe estar claramente identificado. Además, recalca la importancia de formar a los periodistas en competencias tecnológicas para que puedan adaptarse a los cambios sin perder la calidad y los valores del periodismo tradicional.

4.3. Discusión

La integración de la IA en RTVE ilustra un compromiso claro con los principios de ética y transparencia, aunque también plantea desafíos, que van desde la regulación del uso de la tecnología al conocimiento y la percepción pública que se tiene de ella. Las opiniones de Corral y García destacan la importancia de encontrar un equilibrio entre el progreso y la innovación con la responsabilidad ética para, de esta manera, fortalecer el papel de RTVE como medio público de referencia nacional e internacional.

En el ámbito ético, la propuesta de García para desarrollar un código deontológico interno es particularmente relevante. Esta iniciativa podría no solo mejorar la autorregulación de RTVE, sino también podría darle una posición de modelo a seguir en la industria de los medios de

comunicación. Además, la idea de integrar principios de tecnoética en la formación de los periodistas refuerza la necesidad de una alfabetización digital que incluya competencias éticas que serán imprescindibles en los años venideros.

En cuanto a la regulación, la tensión entre la innovación y el control normativo sigue siendo un desafío. Corral enfatiza que Europa debe encontrar un equilibrio que permita competir con Estados Unidos y China sin comprometer los derechos fundamentales. Este dilema subraya la importancia de que RTVE continúe adoptando un enfoque proactivo para cumplir con las normativas mientras lidera la creación de marcos éticos internos.

Finalmente, la lucha contra la desinformación emerge como una prioridad tanto para Corral como para García. Aunque la IA ofrece herramientas poderosas para abordar este fenómeno, su éxito dependerá de una implementación responsable y de la colaboración entre diferentes perfiles profesionales. En este sentido, la estructura multidisciplinar de RTVE es una fortaleza clave que debe ser aprovechada y ampliada.

En definitiva, la integración de la IA en RTVE representa un avance significativo hacia un modelo de medios públicos más eficiente y transparente. Sin embargo, el éxito de esta transición dependerá de la capacidad de la corporación para equilibrar la innovación con los principios éticos y la regulación, consolidándose como un referente en el uso responsable de la IA en los medios de comunicación.

Conclusiones

La integración de la IA en Radio Televisión Española (RTVE) representa un ejemplo de cómo un medio público puede liderar la adaptación tecnológica sin poner en compromiso los valores éticos inherentes a su posición de servicio a la sociedad. En este sentido, RTVE se posiciona como un actor protagonista tanto en la implementación de tecnologías emergentes, como en la promoción de un modelo que conjugue innovación y responsabilidad ética.

El trabajo que se desarrolla por la corporación en la aplicación de la IA refleja una visión integradora que combina el cumplimiento de normativas europeas, como la Ley de IA, con la creación de marcos deontológicos internos orientados a garantizar la transparencia y la confianza del público. Esta doble visión, resaltada tanto por David Corral como por Silvia García en las entrevistas realizadas, refuerza la necesidad de adoptar una perspectiva global que considere tanto las necesidades regulatorias como las especificidades éticas y culturales del entorno mediático español.

Uno de los principales logros, y a la vez retos, de RTVE es el uso de la IA como herramienta para combatir la desinformación, una amenaza creciente en el ecosistema digital actual. A través de la generación automática de noticias, la verificación de datos y la automatización de tareas como el subtítulo de programas, la IA ha permitido optimizar los procesos productivos sin menoscabar la calidad informativa. Este esfuerzo, sin embargo, debe ser complementado por una alfabetización digital de la audiencia y de los propios profesionales, asegurando que la tecnología se utilice como un medio para fortalecer la confianza pública y no como un factor que agrave las desigualdades informativas. Al mismo tiempo, debe establecerse como una herramienta útil para el periodista, y no como un sustituto de este.

La creación de un código deontológico interno, como propone Silvia García, representa una estrategia crucial para consolidar el compromiso ético de RTVE. Este instrumento no solo serviría para regular el uso interno de la IA, sino también podría convertirse en un referente para otros medios de comunicación, tanto a nivel nacional como internacional. En un contexto donde la regulación global de la IA sigue siendo fragmentaria y desigual, RTVE intenta posicionarse como un líder en el debate sobre la ética y la transparencia en el uso de tecnologías emergentes, aunque no es la única televisión pública que avanza en este campo.

Fieiras-Ceide *et al.* (2022) mencionan que en Europa hay corporaciones como la BBC (Reino Unido), ARD (Alemania) y YLE (Finlandia) que presentan desde hace años innovaciones destacadas para mejorar la personalización, la accesibilidad y la verificación de contenidos, como el uso de «medios basados en objetos» para adaptar contenidos según preferencias del usuario, la generación de metadatos y la automatización de tareas o la gestión administrativa y operativa.

A nivel ético, mientras, por ejemplo, France Télévisions (Francia) ha promovido la alfabetización tecnológica entre sus empleados para facilitar la integración de estas herramientas, la BBC ha implementado una normativa detallada para garantizar el uso responsable y ético de la IA en todas sus áreas de creación, distribución y presentación de contenido, para preservar valores como imparcialidad, precisión, transparencia y respeto a la privacidad. El objetivo es que la IA no comprometa la confianza del público ni la calidad del contenido. Los riesgos incluyen sesgos algorítmicos, alucinaciones y plagio.

La BBC protege los derechos de creadores y público, impulsando el equilibrio entre tecnología y la creatividad de los profesionales. La supervisión humana es obligatoria en todas las aplicaciones de IA, y debe ser explicada al público. Los sistemas de terceros requieren autorización previa. Además, el *AI Risk Advisory Group* evalúa los diferentes riesgos legales, éticos y editoriales, reflejando el compromiso de la BBC con la responsabilidad tecnológica y sus valores de servicio público (BBC, 2024).

Otro aspecto destacable es la importancia que RTVE otorga a la colaboración multidisciplinar. La diversidad de perfiles profesionales involucrados en los proyectos de IA subraya la necesidad de enfoques integrados que combinen competencias tecnológicas, periodísticas y éticas. Esta sinergia no solo mejora la calidad de los resultados obtenidos, sino que también asegura una implementación más equitativa y sostenible de la tecnología. Para ello, es prioritaria la actualización de los profesionales y la formación y educación en estas nuevas tecnologías.

En conclusión, la experiencia de RTVE en la adopción de la IA refleja un modelo viable para otros medios de comunicación públicos y privados que deseen integrar estas tecnologías sin comprometer los principios de ética y transparencia. La combinación de normativas estrictas, autorregulación y formación profesional posiciona al ente público como un referente en la industria. Sin embargo, el reto continúa siendo la consolidación de un equilibrio sostenible entre innovación y responsabilidad ética, un objetivo que no solo determinará el futuro de RTVE, sino también el papel de la IA en la configuración del periodismo del siglo XXI.

Referencias

- Aguilera, P. P. (2023). La inteligencia artificial como «sujeto» ético del periodismo actual. En: Gutiérrez Atala, F. y Muñoz Joven, L. A. (eds.): *Ética y moralidad en los medios de comunicación: investigaciones y propuestas* (281-303). Universidad Santiago de Cali. <https://bit.ly/4gnxX7J>.
- AI Safety Summit. (2023). «The Bletchley Declaration». GOV.UK. de 2023 <http://u.uma.es/fWw/bletchley/>.
- BBC. (2024). *Guidance: The use of Artificial Intelligence*. <https://bit.ly/4aZALqV>.
- Blanco, S., Martín-Martín, F. M. y Sedano, J. (2023). La visibilidad mediática de la desinformación en los programas informativos: El caso de La 1 de RTVE. *Estudios sobre el Mensaje Periodístico*, 29(4), 893-904. <https://doi.org/10.5209/esmp.88595>.
- Fieiras Ceide, C., Vaz Álvarez, M. y Túñez López, J. M. (2022). Artificial intelligence strategies in European public broadcasters: Uses, forecasts and future challenges. *El Profesional de La Información*, 31(5), e310518. <https://doi.org/10.3145/epi.2022.sep.18>.
- Gómez-Calderón, B. y Ceballos, Y. (2024). Periodismo e inteligencia artificial. El tratamiento de los chatbots en la prensa española. *index.comunicación*, 14(1). <https://doi.org/10.62008/ixc/14/01Period>.
- González Esteban, E. y Sanahuja Sanahuja, R. (2023). Exigencias éticas para un periodismo responsable en el contexto de la inteligencia artificial. *Daimon Revista Internacional de Filosofía*, 90. <https://doi.org/10.6018/daimon.557391>.

- Guallar, J., Codina Bonilla, L., Freixa Font, P. y Pérez-Montoro Gutiérrez, M. (2020). Desinformación, bulos, curación y verificación: Revisión de estudios en Iberoamérica 2017-2020. *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 22(3), 595-613. <https://doi.org/10.36390/telos223.09>.
- Lopezosa, C., Codina Bonilla, L., Pont i Sorribes, C. y Vallez, M. (2023). Use of generative artificial intelligence in the training of journalists: Challenges, uses and training proposal. *El Profesional de La Información*, 32(4), 1-12. <https://doi.org/10.3145/epi.2023.jul.08>.
- Martín-Crespo Blanco, M. C. y Salamanca Castro, A. B. (2007). El muestreo en la investigación cualitativa. *NURE Investigación*, (27). <https://bit.ly/4ienQEo>.
- Mayoral Sánchez, J., Parratt Fernández, S. y Mera Fernández, M. (2023). Uso periodístico de la IA en medios de comunicación españoles: Mapa actual y perspectivas para un futuro inmediato. *Estudios sobre el Mensaje Periodístico*, 29, 821-832. <https://doi.org/10.5209/esmp.89193>.
- Montiel Torres, M. F., Zafra Arroyo, A. y Teruel Rodríguez, L. (2024). Confianza en los medios de comunicación y libertad de prensa en el sur de Europa: Diferencias entre países y perfiles sociodemográficos. *Estudios sobre el Mensaje Periodístico*, 30(3), 553-568. <https://doi.org/10.5209/emp.96323>.
- Palomo, B., Tandoc, E. C. y Cunha, R. (2023). El impacto de la desinformación en las rutinas profesionales y soluciones basadas en la inteligencia artificial. *Estudios sobre el Mensaje Periodístico*, 29(4), 757-759. <https://doi.org/10.5209/esmp.92160>.
- Parlamento Europeo. (2024). *Acta sobre Inteligencia Artificial (TA-9-2024-0138)*. <https://bit.ly/40GwyDt>.
- Parrat-Fernández, S., Chaparro-Domínguez, M. Á. y Martín-Sánchez, I. M. (2024). Cobertura mediática de la inteligencia artificial periodística en España: Relevancia, temas y framing. *Revista mediterránea de Comunicación*, 15(2), e25169. <https://www.doi.org/10.14198/MEDCOM.25169>.
- Quian, A. y Sixto García, J. (2024). Inteligencia artificial en la prensa: Estudio comparativo y exploración de noticias con ChatGPT en un medio tradicional y otro nativo digital. *Revista de Comunicación*, 23(1), 457-483. <https://doi.org/10.26441/RC23.1-2024-3374>.
- Quimis Arteaga, L. E. (2024). Inteligencia Artificial en el Periodismo: Innovación, Ética y colaboración en la creación y verificación de contenidos. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(6), 1120-1139. <https://doi.org/10.56712/latam.v5i6.3069>.
- Real Rodríguez, E., Príncipe Hermoso, S. y Agudiez Calvo, P. (2024). La transformación digital de la televisión pública. Estudio de caso de RTVE, Rai y RTP. *Estudios sobre el Mensaje Periodístico*, 30(1), 211-221. <https://doi.org/10.5209/esmp.91920>.
- Russell, S. y Norvig, P. (1995). *Artificial Intelligence: Modern Approach* (4.a ed.). Prentice Hall.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460. <https://doi.org/10.1093/mind/lix.236.433>.

- Valencia Bermúdez, A. y Túnnez López, J. M. (2018). La comunicación del valor social del servicio audiovisual público. En: López Golán, M., Campos Freire, F., López López, P. C. y Rivas Echeverría, F. (eds.): *La comunicación en la nueva sociedad digital* (293-319). Universidad de Los Andes (ULA). <https://bit.ly/4aNQ19Z>.
- Vargues-Vara, D. (2024). Las nuevas reglas del juego que la Inteligencia Artificial introduce en el periodismo. *Revista Mediterránea de Comunicación*, 15(2), e26771. <https://doi.org/10.14198/MEDCOM.26771>.
- Villalobos Portalés, J. (2023). La propuesta comunitaria de ley de inteligencia artificial y el «invierno legislativo». *Desafíos Jurídicos*, 3(5), 44-57. <https://doi.org/10.29105/dj3.5-114>.

Humano ex machina: enfoque y ejemplos para la creación de herramientas de inteligencia artificial contra la desinformación

Javier Cantón-Correa

Universidad de Granada / Universidad Internacional de La Rioja
javicanton@ugr.es / javier.canton@unir.net

It's easier to fool people than to convince them
that they have been fooled.
Mark Twain

En pleno auge de la inteligencia artificial aplicada a múltiples áreas de la vida humana, en el imaginario colectivo yace la idea de que su uso deriva en automatizaciones capaces de realizar tareas muy cercanas a las propias capacidades humanas, sustituyéndonos en aquellas especialmente repetitivas o farragosas. En ámbitos como la lucha contra la desinformación, esto se traduciría en la posibilidad de crear «agentes artificiales» a los que les pudiéramos suministrar una información dudosa para que evaluaran, etiquetaran y catalogaran dicha información como falsa o no de una manera rápida. Pero la realidad dista de tal simplicidad tecnológica. Si a nosotros mismos nos cuesta muchas veces discernir los elementos de verosimilitud en el lenguaje humano (no digamos ya en las imágenes), compréndase la dificultad de entrenar a una máquina para realizar esta tarea compleja. Por eso, en los últimos años se apuesta por un enfoque más amplio donde las tecnologías de IA no

son un sustituto, sino un aliado, un ayudante, al ser humano encargado de dicha labor: es lo que se denomina el enfoque «*human-in-the-loop*» o «humano en el bucle», el diseño de un sistema asistido de IA para la realización de tareas complejas donde la máquina no tiene la última palabra, sino el ser humano.

1 | Desinformación, no «fake news»

El fenómeno de la desinformación no es nuevo en la historia humana, pero ha adquirido nuevas formas en la era digital. La posverdad, entendida como un entorno donde los hechos objetivos tienen menos influencia en la opinión pública que las emociones y creencias personales, ha permitido la proliferación de información manipulada con fines políticos, económicos o ideológicos (McIntyre, 2018). Según Claire Wardle (2017), en lugar de noticias falsas o «*fake news*» es mejor hablar de desórdenes de la información. Por un lado, porque el término ha sido desvirtuado y politizado, siendo usado para desacreditar información incómoda más que para referirse a contenido realmente falso y por el otro, porque es el término es un oxímoron: por definición, una noticia periodística, contrastada, no puede ser falsa. Estos contenidos falsos encuadrados en los desórdenes de la información pueden clasificarse en *misinformation* o información errónea (información errónea sin intención de manipulación), *malinformation* o información improcedente (información real usada con fines perjudiciales) y *disinformation* o información engañosa (información falsa difundida con intención de engañar), que comúnmente conocemos como en español como «desinformación» (Wardle, 2017). Por tanto, la desinformación se enmarca dentro de estos desórdenes de la información y se define como la difusión intencionada de información falsa con el objetivo de manipular o influir en la opinión pública. A diferencia de la *misinformation*, que hace referencia a información errónea compartida sin intención de engañar, y la *malinformation*, que implica el uso de información veraz con el fin de causar daño, la desinformación se caracteriza por su intencionalidad en la alteración de la percepción de la realidad (Cantón-Correa, 2023) perteneciente al Máster de Inteligencia Artificial de la Universidad Internacional de La Rioja (UNIR) (fig. 8.1).

La aceleración y amplificación de estos desórdenes informativos en el ecosistema digital actual se debe, en gran medida, a la transformación de las dinámicas de consumo y distribución de la información. Las redes sociales y plataformas de mensajería han reemplazado en buena medida a los medios tradicionales como principales fuentes de información, facilitando la viralización de contenidos sin los filtros editoriales y de verificación previos. Además, la estructura algorítmica de estas plataformas promueve la personalización extrema del contenido, creando burbujas

Types of Information Disorders

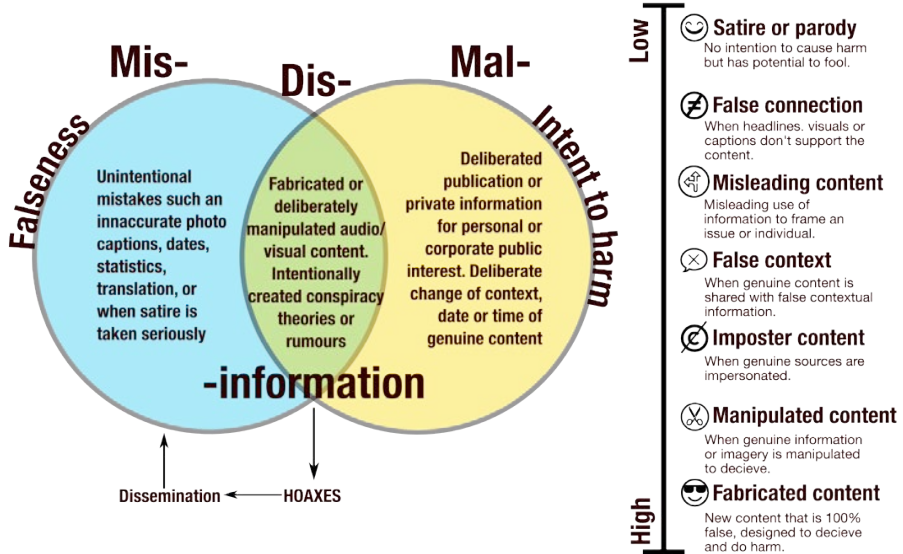


Figura 8.1. Tipos de desórdenes informativos. Fuente: elaboración propia a partir de Wardle (2017)

informativas donde los usuarios son expuestos mayoritariamente a información que refuerza sus creencias previas. Esto genera entornos propicios para la propagación de desinformación, ya que el contenido que apela a la emoción o la indignación tiene más probabilidades de ser compartido y alcanzar grandes audiencias, independientemente de su veracidad. En este contexto, la desinformación no solo afecta el ámbito político o social, sino que se extiende a temas como la salud pública, la ciencia o la economía, con consecuencias directas en la percepción y toma de decisiones de la ciudadanía (Magallón, 2019).

La amplificación de la desinformación en entornos digitales no solo responde a la estructura algorítmica de las plataformas, sino también a sesgos cognitivos que afectan cómo las personas procesan y difunden información. El sesgo de confirmación refuerza burbujas informativas al hacer que los usuarios busquen y compartan contenido alineado con sus creencias previas, mientras que el efecto Dunning-Kruger hace que personas con escaso conocimiento sobre un tema sobrestimen su capacidad de discernir información veraz, facilitando la propagación de errores. Además, el sesgo de disponibilidad favorece la credibilidad de contenidos recientes o impactantes sin una verificación rigurosa, y el efecto de ilusión de verdad hace que la repetición constante de una afirmación aumente su percepción de veracidad, independientemente de su exactitud. Estos sesgos, junto con la lógica de las plataformas digitales, potencian

la viralización de la desinformación y su impacto en la toma de decisiones sociales y políticas.

Estos sesgos, combinados con la velocidad y escala de propagación que permiten las plataformas digitales, explican en gran medida por qué la desinformación es tan difícil de combatir y por qué la mera disponibilidad de información verificada no siempre es suficiente para contrarrestarla. En este contexto, las herramientas de IA pueden desempeñar un papel fundamental en la detección y mitigación de estos contenidos, pero su uso también enfrenta desafíos importantes, como veremos después.

1.1. La respuesta: el periodismo de verificación

Como respuesta a la crisis de credibilidad que provoca el aumento en este tipo de desórdenes informativos, surge el periodismo de verificación. En este contexto, el *fact-checking* se ha convertido en una disciplina clave dentro del periodismo, especializada en la comprobación de hechos mediante métodos tradicionales de investigación y herramientas digitales avanzadas.

El periodismo de verificación y el *fact-checking* (la parte de la verificación dedicada especialmente a verificar afirmaciones realizadas por actores políticos) son ramas del periodismo especializadas en la comprobación de la veracidad de afirmaciones públicas, noticias y otros contenidos que circulan en medios digitales y tradicionales. No es nada más (y nada menos) que desgajar una parte integral del propio proceso periodístico, el contraste y la validación de la información, a una especialización laboral del campo de la información, precisamente debido al aumento exponencial de la información que circula por nuestros ecosistemas informativos y, por tanto, al auge de estos desórdenes informativos, en cantidad y calidad.

Este tipo de periodismo ha ganado relevancia en un contexto de *infoxicación*, donde la sobrecarga de información dificulta discernir lo verdadero de lo manipulado o descontextualizado. Como afirman Ufarte-Ruiz *et al.* (2018), la verificación de información representa tanto una serie de nuevas competencias necesarias para los periodistas como una especialización dentro del ámbito periodístico. La verificación no solo se basa en técnicas tradicionales de investigación periodística, sino que también incorpora métodos computacionales avanzados para hacer frente al volumen y la velocidad de propagación de la desinformación. Graves *et al.* (2016) definen este proceso como la comprobación de hechos mediante el uso de fuentes primarias y secundarias, contrastando datos con registros oficiales, expertos y otras referencias verificadas. A pesar del avance de herramientas automatizadas, la verificación de información sigue dependiendo en gran medida del juicio humano, ya que la interpretación

del contexto, la intención del mensaje y la identificación de sesgos requieren un análisis crítico que las máquinas aún no pueden replicar completamente. En este sentido, el papel de los verificadores de hechos es clave para garantizar la calidad informativa y combatir la proliferación de bulos en la esfera pública. De hecho, Graves (2018) también destaca cómo los verificadores han evolucionado de meros observadores a actores clave en la lucha contra la desinformación, con iniciativas en España como Newtral o Maldita.es, que utilizan modelos de IA para la detección y verificación de información.

2 | **Inteligencia artificial para crear desinformación y para luchar contra ella**

El uso de la IA en la generación y difusión de desinformación ha crecido exponencialmente en los últimos años, facilitado por el desarrollo de modelos avanzados de generación de texto, imágenes y videos sintéticos. Herramientas basadas en modelos de lenguaje como GPT-4 han demostrado una capacidad sin precedentes para generar contenido textual convincente y difícil de distinguir de textos escritos por humanos (Zellers *et al.*, 2019). Esto ha permitido la automatización de campañas de desinformación a gran escala, donde actores malintencionados pueden crear artículos falsos, comentarios automatizados y narrativas manipuladas en redes sociales con gran rapidez y volumen. El uso de bots y granjas de contenidos impulsados por IA permite la amplificación de estos mensajes en plataformas digitales, a través de campañas de *astroturfing*, aumentando su visibilidad y credibilidad percibida entre los usuarios (Marín *et al.*, 2023).

Uno de los desarrollos más preocupantes en este ámbito es el uso de *deepfakes* o «ultrafalsos», contenido audiovisual manipulado mediante redes neuronales generativas adversarias (GANs) para alterar la apariencia y voz de personas reales en videos falsos altamente realistas (Chesney & Citron, 2019). Estas tecnologías han sido utilizadas para fabricar discursos políticos inexistentes, crear pruebas falsas en contextos judiciales e incluso manipular eventos históricos con fines propagandísticos. La proliferación de *deepfakes* plantea desafíos significativos para la verificación de la información, ya que los métodos tradicionales de detección de manipulación audiovisual se vuelven obsoletos frente a la creciente sofisticación de estos modelos (Vaccari & Chadwick, 2020).

Además de la generación de contenido falso, la IA también se emplea para optimizar la difusión de la desinformación mediante técnicas avanzadas de segmentación y *microtargeting* en redes sociales. Plataformas como Facebook y Twitter han utilizado durante años algoritmos de recomendación que priorizan contenido altamente emocional y polarizante, lo que facilita la propagación de información engañosa entre comunidades

específicas. Los mismos modelos que optimizan la publicidad personalizada se han adaptado para potenciar campañas de desinformación dirigidas a audiencias vulnerables, explotando sesgos cognitivos como el de confirmación y reforzando burbujas informativas. Este fenómeno ha llevado a debates sobre la necesidad de regular el uso de la IA en la manipulación informativa y la responsabilidad de las plataformas digitales en la moderación de contenido falso (Hao, 2021).

Pero la IA también se está convirtiendo en una herramienta fundamental en la **lucha contra la desinformación**. Sin embargo, al ser la desinformación un fenómeno complejo, con múltiples dimensiones (intencionalidad del contenido, contexto, forma en que se difunde...), no es abordable con un único enfoque. Ninguna herramienta es capaz de resolver por sí sola el problema de la desinformación (Montoro-Montarroso *et al.*, 2023). Los sistemas más efectivos integran varias soluciones en cada fase del flujo de trabajo de los verificadores, combinando la detección automatizada de afirmaciones sospechosas con la validación mediante bases de datos de hechos contrastados, el rastreo de evidencias en la web y la clasificación del contenido según su grado de veracidad. Estas soluciones utilizan técnicas avanzadas como el aprendizaje automático, el procesamiento del lenguaje natural y el análisis de redes sociales para identificar patrones en la creación y la propagación de información falsa.

- El *machine learning* o **aprendizaje automático** ha demostrado ser una de las técnicas más efectivas en la lucha contra la desinformación, al permitir la identificación de patrones en grandes volúmenes de datos. Mediante modelos de aprendizaje supervisado y no supervisado, la IA puede clasificar contenidos, analizando sus características lingüísticas, semánticas y contextuales. Técnicas como las máquinas de soporte vectorial (SVM), redes neuronales y árboles de decisión han sido ampliamente utilizadas para entrenar modelos que tratan de identificar la desinformación entrenándose a partir de bases de datos previamente etiquetadas por verificadores. Sin embargo, la efectividad de estos modelos depende precisamente de la calidad y diversidad de los datos con los que se entrenan, lo que plantea el desafío de actualizar constantemente las bases de datos utilizadas para la clasificación (Montoro-Montarroso *et al.*, 2023).
- Por otro lado, el *Natural Language Processing* o **procesamiento del lenguaje natural** permite analizar textos a nivel sintáctico y semántico, proporcionando información valiosa para detectar contenido engañoso. Esta técnica se basa en la extracción de características estilísticas de los textos, como la frecuencia de ciertas estructuras gramaticales, el uso de adjetivos emocionales o la presencia de patrones lingüísticos que suelen encontrarse en textos desinformativos. Algoritmos como *word embeddings* y

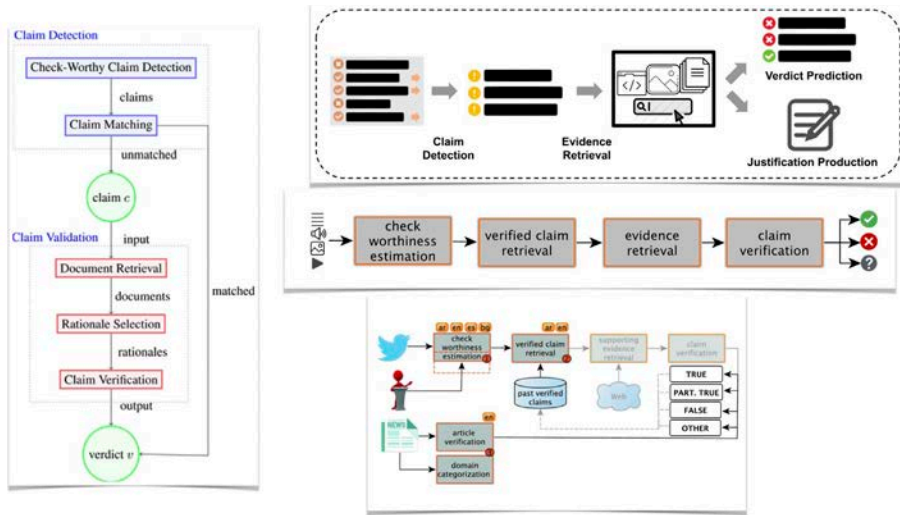


Figura 8.2. Distintos modelos de aproximación al desarrollo de herramientas de IA contra la desinformación. Fuente: elaboración propia

modelos de redes neuronales han facilitado la clasificación automática de contenido falso al identificar similitudes en el discurso utilizado en noticias engañosas. Un ejemplo de aplicación de PLN en la verificación de información es ClaimCheck, una herramienta creada por Newtral (2022) que analiza publicaciones en redes sociales y las compara con bases de datos de afirmaciones previamente verificadas para alertar sobre contenidos potencialmente falsos (Cantón-Correa, 2023) perteneciente al Máster de Inteligencia Artificial de la Universidad Internacional de La Rioja (UNIR).

- El *Social Network Analysis* o **análisis de redes sociales** ha sido clave para rastrear la propagación de la desinformación, identificando patrones en la difusión de contenido falso. Esta técnica permite detectar comunidades de usuarios que amplifican narrativas engañosas, así como el uso de cuentas automatizadas (bots) que generan tráfico artificial para aumentar la visibilidad de ciertas noticias. A través del ARS, es posible mapear redes de desinformación y analizar cómo ciertos actores impulsan mensajes con intenciones manipulativas. Un ejemplo de su aplicación es la identificación de campañas de desinformación coordinadas en Twitter, donde ciertos hashtags se viralizan artificialmente para posicionar narrativas específicas en la conversación pública (Cantón-Correa, 2023) perteneciente al Máster de Inteligencia Artificial de la Universidad Internacional de La Rioja (UNIR, fenómeno conocido como *astroturfing*) (fig. 8.2).

3 | El enfoque *human-in-the-loop* en la creación de herramientas de inteligencia artificial

Así pues, las tecnologías de IA han permitido la creación de herramientas que asisten a los verificadores en diferentes etapas del proceso de verificación. Ejemplos como ClaimCheck de Newtral (2022) han demostrado que la IA puede identificar patrones lingüísticos en afirmaciones verificadas previamente y alertar sobre posibles repeticiones de desinformación. Sin embargo, la confianza en estos sistemas sigue siendo un desafío, ya que, dada la imposibilidad de una automatización completa en la verificación de información, los sistemas híbridos están ganando relevancia para los verificadores: así, la IA complementa, pero no sustituye, el juicio humano (Nakov *et al.*, 2021).

Es lo que se conoce como enfoque *human-in-the-loop*, que combina el poder analítico de la IA con la capacidad de discernimiento humano, permitiendo que los verificadores ajusten y refinen las predicciones algorítmicas en tiempo real (Cantón-Correa, 2023; La Barbera *et al.*, 2022) perteneciente al Máster de Inteligencia Artificial de la Universidad Internacional de La Rioja (UNIR). Este modelo no solo reduce el impacto de los sesgos algorítmicos, sino que también mejora la confianza en las herramientas de verificación al garantizar que la toma de decisiones sigue dependiendo del criterio experto. Este enfoque también permite una retroalimentación constante entre el humano y el sistema, lo que facilita la mejora progresiva de los modelos de IA. A medida que los verificadores etiquetan y corrigen predicciones erróneas, los algoritmos pueden ajustar sus parámetros y mejorar su precisión en futuras evaluaciones. Este ciclo de mejora es fundamental para abordar la evolución de la desinformación, ya que los discursos y estrategias de manipulación cambian constantemente, requiriendo sistemas adaptativos que no dependan únicamente de bases de datos preexistentes.

Otra ventaja clave del enfoque *human-in-the-loop* es su capacidad para interpretar el contexto, un aspecto donde la IA aún presenta limitaciones significativas. Las herramientas automatizadas pueden identificar patrones sospechosos y detectar incoherencias en una afirmación, pero carecen de la comprensión profunda del significado, la ironía o la intención detrás del mensaje. En este sentido, los verificadores humanos pueden aportar un análisis contextualizado, diferenciando entre una noticia satírica y una desinformación maliciosa, o identificando matices en declaraciones ambiguas. Algo que aún no pueden hacer las máquinas (fig. 8.3).

Los autores de este capítulo trabajamos en el desarrollo de herramientas con este enfoque *human-in-the-loop*, combinando la capacidad de análisis automatizado de la inteligencia artificial con la supervisión y el juicio humano. Buscamos abordar distintos aspectos del problema, desde la monitorización de redes hasta la detección de patrones lingüísticos en narrativas desinformativas. A continuación, vamos a presentar

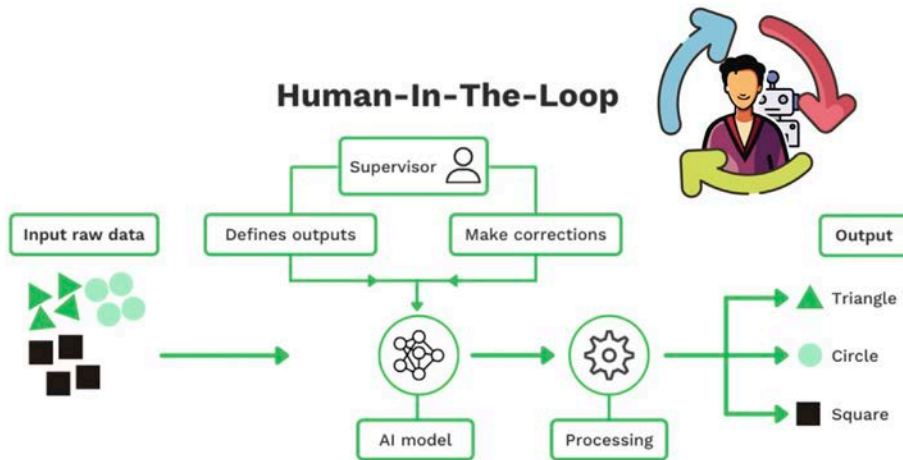


Figura 8.3. Diagrama explicativo del paradigma *human-in-the-loop*. Fuente: <https://www.klippa.com/es/blog/informativo/human-in-the-loop-automatizar/>

tres de las herramientas que hemos desarrollado bajo este enfoque: **SoundHive**, una plataforma de monitorización de Telegram; **pytopicgram**, un sistema de análisis temático, y un modelo de **detección de mensajes conspirativos en Telegram**, cada una de ellas con aplicaciones específicas para la verificación de información y la mitigación de la desinformación.

Nótese que nos centramos en Telegram como eje del desarrollo de nuestras herramientas debido a su creciente papel en la difusión de desinformación y su relativa falta de control en comparación con otras plataformas (Cantón-Correa, 2023) perteneciente al Máster de Inteligencia Artificial de la Universidad Internacional de La Rioja (UNIR). A diferencia de otras plataformas, Telegram permite la creación de grupos y canales con miles de miembros donde la información circula sin prácticamente restricciones. Su arquitectura descentralizada, combinada con el anonimato que ofrece a los usuarios, facilita la propagación de narrativas falsas, campañas de manipulación y discursos conspirativos sin una supervisión efectiva. Además, estudios recientes han señalado que Telegram se ha convertido en un espacio clave para la coordinación de estrategias desinformativas, donde los mensajes pueden viralizarse rápidamente y luego ser redistribuidos en otras plataformas (Maldita.es, 2024), especialmente en aquellas como X y Facebook, donde se han desmantelado los equipos existentes encargados de la moderación de contenidos (Cantón-Correa, 2025). Por ello, resulta fundamental desarrollar herramientas específicas que permitan monitorizar y analizar el flujo de información dentro de este entorno, proporcionando a los verificadores una visión más clara de cómo evolucionan las narrativas engañosas en tiempo real y permitiéndoles anticipar su impacto en el debate público.

3.1. Monitorización de desinformación

La práctica de la verificación requiere de una monitorización constante de las fuentes de información, en búsqueda de contenidos que se hacen pasar por contrastados siendo falsos. Cada día, muchos verificadores comienzan su jornada de trabajo comprobando distintas listas y canales donde habitualmente se difunden contenidos desinformativos de manera constante. Herramientas que faciliten esta búsqueda y ayuden en este proceso de monitorización son cada vez más útiles y necesarias en esta labor.

En ese sentido, la herramienta creada llamada *SoundHive* es un sistema diseñado para la monitorización en tiempo real de mensajes en Telegram. Se inicia a partir de un trabajo más amplio (Cantón-Correa, 2023) perteneciente al Máster de Inteligencia Artificial de la Universidad Internacional de La Rioja (UNIR, que planteaba un sistema de monitorización a mayor escala, centrado en «escuchar» distintas fuentes de información, concebido de manera modular, de la que se desgajó y desarrolló el módulo relativo a Telegram, gracias a la facilidad y gratuidad del uso de la API¹ de Telegram. Esta misma idea de construcción modular, en este caso aplicada a la detección de desinformación en plataformas y redes sociales, es aplicable también a otras temáticas, como el análisis de tendencias a tiempo real a partir de esta misma «escucha» en una selección de fuentes, por ejemplo, monitorizar y detectar automáticamente noticias mediante IA (Cantón-Correa *et al.*, 2024).

La herramienta prototipada, que actualmente se encuentra en desarrollo a través de un proyecto de transferencia², permitiría a los verificadores de información analizar grandes volúmenes de mensajes de canales públicas de Telegram, identificar tendencias emergentes y detectar posibles patrones en la propagación de contenidos falsos. La idea de fondo es practicar lo que se conoce como *prebunking*, esto es, en lugar verificar un contenido cuando se ha hecho viral (*debunking*), tratar de adelantarse a dicha viralización y detectarlo previamente para atajar el contenido lo antes posible. Esto implica ser capaces de evaluar el alcance de ciertas narrativas antes de que se viralicen, para lo que se utilizan técnicas de IA.

1. Una API (*Application Programming Interface* o Interfaz de Programación de Aplicaciones) es un conjunto de funciones y protocolos que permiten a los desarrolladores interactuar con una plataforma o servicio de manera automatizada. En el caso de Telegram, su API proporciona acceso a diversas funcionalidades de la aplicación, como la recuperación de mensajes de canales y grupos públicos, la gestión de bots y la extracción de metadatos. Esto permite a herramientas como la mencionada recopilar y analizar contenido en tiempo real de manera automatizada.

2. *MonitorIA: herramienta avanzada para la monitorización y detección temprana de desinformación en Telegram mediante IA* (T25-003-I), 2025-26. Proyecto de transferencia UNIR, convocatoria Incentiva 25.

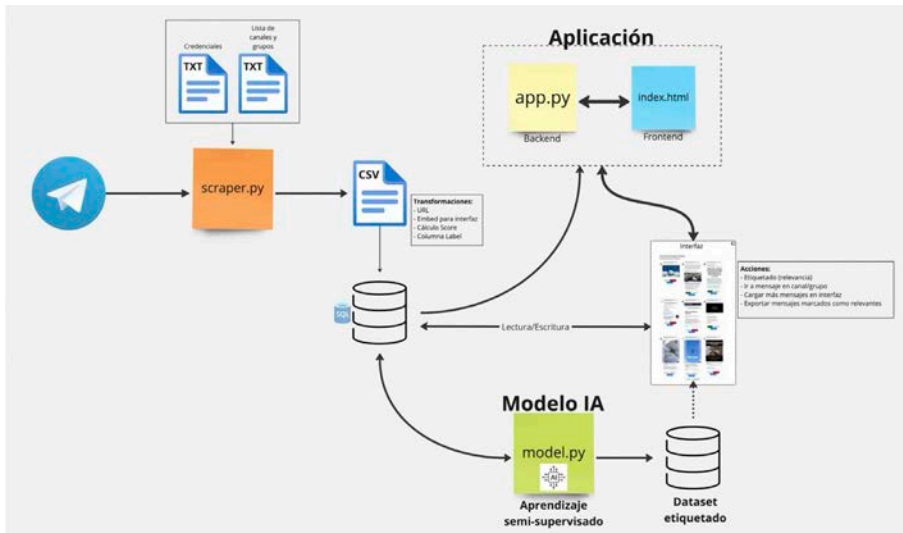


Figura 8.4. Diagrama explicativo del funcionamiento de la herramienta: los datos se extraen desde Telegram, se cargan en una base de datos de la que se alimenta la aplicación, que muestra al usuario una interfaz con la que interactuar con la información y entrenar el modelo de IA para detectar desinformación. Fuente: elaboración propia (Cantón-Correa, 2023)

Las técnicas de IA usadas son el procesamiento de lenguaje natural (PLN) y el aprendizaje semisupervisado. Por un lado, el PLN permite analizar el contenido textual de los mensajes, identificando estructuras lingüísticas y términos que pueden aparecer en contextos desinformativos (muchas narrativas desinformadoras usan códigos y lenguajes propios para tratar de eludir una posible censura). Esto incluye el uso de palabras clave específicas, la detección de frases emocionalmente cargadas y la identificación de patrones argumentativos comunes en bulos y teorías conspiracionistas.

Por otro lado, el aprendizaje semisupervisado permite que el sistema refine su precisión a lo largo del tiempo mediante la retroalimentación de los verificadores humanos: la interfaz ofrece una selección de los mensajes con mayor impacto en las respectivas audiencias de los canales y permite al verificador elegir si cada mensaje es relevante o no para el modelo, que aprende de este entrenamiento. A medida que los especialistas etiquetan y validan mensajes, el modelo ajusta sus parámetros para mejorar su capacidad de detección y reducir los falsos positivos, tratando de llegar a un equilibrio entre automatización y supervisión humana. Cuando detecta un mensaje cuyo análisis muestra una susceptibilidad de ser catalogado como «contenido falso» marca el mensaje con una señal de alerta.

En términos prácticos, la herramienta planteada podría ser de gran utilidad para periodistas, investigadores y organizaciones dedicadas a la verificación de información. Su implementación facilitaría esta detección temprana de bulos y campañas de desinformación organizadas y permitiría que los verificadores actúen de manera preventiva, antes de que estos contenidos alcancen una difusión masiva. Al proporcionar un sistema de alerta sobre la propagación de mensajes engañosos, esta herramienta optimizaría también los tiempos de respuesta y contribuiría a desarrollar estrategias de comunicación más efectivas para contrarrestar la desinformación en Telegram.

3.2. Análisis temático de narrativas desinformativas

Otra herramienta que muestra las posibilidades de las técnicas de IA aplicadas a la lucha contra la desinformación es *pytopicgram*, diseñada para detectar y analizar la evolución de narrativas desinformativas en redes sociales (Gómez-Romero, Cantón-Correa *et al.*, 2024). En lugar de centrarse en la verificación de hechos puntuales, este sistema permite estudiar cómo ciertos discursos falsos o engañosos se transforman y adaptan con el tiempo. La idea de fondo es, a través de este enfoque, identificar patrones emergentes en la desinformación y anticiparse a su viralización. En primer lugar, porque la verificación de información cada vez se centra más en narrativas y menos en bulos concretos y, además, porque comprender la trayectoria de una narrativa desinformativa es crucial para desarrollar estrategias de mitigación cada vez más efectivas, ya que muchas campañas de desinformación se reciclan con ligeros cambios para seguir siendo relevantes. Los canales donde se difunde esta desinformación también cambian de narrativas para atraer a más usuarios: donde en 2020 se hablaba contra las vacunas y las medidas estatales para luchar contra la covid, en 2025 se ha virado hacia otros temas actualmente polémicos y polarizantes, como la inmigración, la política nacional o la ocupación de pisos.

Para lograr este análisis, *pytopicgram* utiliza técnicas avanzadas de modelado de temas, como Latent Dirichlet Allocation (LDA) y *embeddings* de palabras. Sin entrar en muchos detalles técnicos, el LDA permite identificar los temas predominantes dentro de grandes conjuntos de datos, clasificando y organizando publicaciones en grupos temáticos similares semánticamente. Esto facilita la detección de tendencias y permite rastrear cómo ciertas ideas evolucionan o se reconfiguran con el tiempo. Por otro lado, los *embeddings* de palabras ayudan a establecer relaciones entre términos clave o *keywords*, identificando conexiones entre discursos aparentemente independientes. Esto es particularmente útil para descubrir estrategias de manipulación donde se vinculan narrativas previamente desacreditadas con nuevos contextos para mantener su impacto (fig. 8.5).

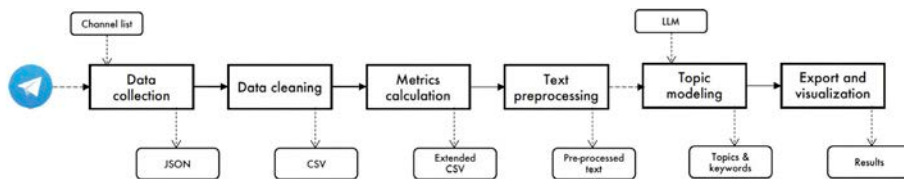


FIGURA 8.5. Diagrama ilustrativo del funcionamiento de la herramienta Pytopicgram: los datos van siendo procesados hasta servir para entrenar el modelo de ia que detecta y muestra los topics o diferentes temas que subyacen en las publicaciones de Telegram. Fuente: elaboración propia (Gómez-Romero, Cantón-Correa, *et al.*, 2024)

La herramienta ya se ha podido usar en investigaciones activas, realizadas en colaboración con entidades de verificación para detectar cómo se mueve la desinformación en canales públicos de Telegram y cómo han evolucionado las narrativas utilizadas en dichos canales. Concretamente, la investigación analizó cerca de 1,5 millones de mensajes en 95 canales públicos de Telegram en español, revelando que estos canales auto-denominados «alternativos» tenían más seguidores y generaban mayor interacción que los medios tradicionales. Estos canales, que difundían desinformación y teorías de la conspiración, formaban comunidades interconectadas y adaptaban sus mensajes a la actualidad, ofreciendo a sus seguidores una visión muy distorsionada y sesgada de los acontecimientos (Maldita.es, 2024).

El hecho de que *pytopicgram* esté disponible en GitHub³ permite que investigadores, verificadores y desarrolladores accedan libremente a su código, lo que facilita su uso, mejora y adaptación a distintos contextos de análisis de desinformación⁴. Además, su código abierto fomenta la transparencia y la colaboración, permitiendo que la comunidad científica y técnica contribuya a optimizar sus modelos y ampliar su aplicabilidad en la detección de narrativas engañosas.

El uso de *pytopicgram* puede proporcionar a investigadores y verificadores una visión amplia de cómo se estructura y propaga la desinformación en el entorno digital. Al analizar la evolución de los discursos falsos, esta herramienta permite diseñar respuestas informativas más efectivas, ya sea a través de campañas de concienciación o la generación de contranarrativas basadas en datos. Además, su aplicación en redes

3. GitHub es una plataforma de desarrollo colaborativo basada en Git, un sistema de control de versiones que permite gestionar y rastrear cambios en el código de *software*. Se utiliza ampliamente en proyectos de código abierto y privados, facilitando la colaboración entre desarrolladores, investigadores y organizaciones. A través de GitHub, los usuarios pueden compartir, mejorar y adaptar herramientas como *pytopicgram*, contribuyendo a su evolución y aplicabilidad en distintos contextos.

4. <https://github.com/ugr-sail/pytopicgram>.

sociales podría facilitar la identificación de comunidades vulnerables a la desinformación, lo que permite dirigir esfuerzos de alfabetización mediática y prevención hacia los sectores más afectados.

3.3. Detección de mensajes conspirativos

El análisis de mensajes conspirativos es una estrategia clave para identificar la desinformación antes de que se convierta en una narrativa dominante, en ese enfoque que ya hemos comentado previamente de *pre-bunking*. Esta última herramienta que presentamos aquí para la detección de mensajes conspirativos en Telegram está diseñada para identificar y analizar discursos desinformativos dentro de la plataforma, diferenciando entre el pensamiento crítico legítimo y las narrativas conspirativas sin fundamento (Gómez-Romero, González-Silot *et al.*, 2024). Para lograrlo, emplea un sistema de clasificación binaria, que permite distinguir los mensajes según su contenido y estructura discursiva. Uno de sus principales avances es la implementación de anonimización de entidades nombradas, un proceso que reemplaza referencias a personas, organizaciones y lugares por etiquetas genéricas. Esto evita que el modelo dependa de términos específicos y mejora su capacidad de generalización, facilitando la detección de patrones lingüísticos propios de la desinformación conspirativa, independientemente del contexto concreto en el que aparezcan.

Desde un punto de vista técnico, la herramienta utiliza redes neuronales de tipo *feed-forward* (modelos de IA en los que la información fluye en una sola dirección, sin ciclos ni retroalimentación) y *embeddings* de oraciones (representaciones matemáticas de frases que capturan su significado y relación con otros textos) para analizar mensajes en inglés y español, logrando un alto grado de precisión en la clasificación⁵. Al enfocarse en la estructura del lenguaje en lugar de en nombres específicos, la herramienta permite detectar narrativas conspirativas que evolucionan o se adaptan con el tiempo, sin requerir una reconfiguración constante.

Esta tecnología tiene una gran utilidad en la lucha contra la desinformación, pues su capacidad de identificar tendencias emergentes en tiempo real permite a verificadores de información y analistas anticiparse a la propagación de bulos y generar respuestas informativas antes de que alcancen una difusión masiva. Además, por su nivel de generalización, es adaptable a múltiples idiomas, lo que amplía su aplicabilidad en monitorización global de la desinformación (fig. 8.6).

5. La evaluación de su desempeño mediante la métrica MCC (Matthews Correlation Coefficient) muestra valores de 0,797 en inglés y 0,672 en español, lo que indica una notable capacidad para identificar contenido desinformativo sin necesidad de intervención humana directa.

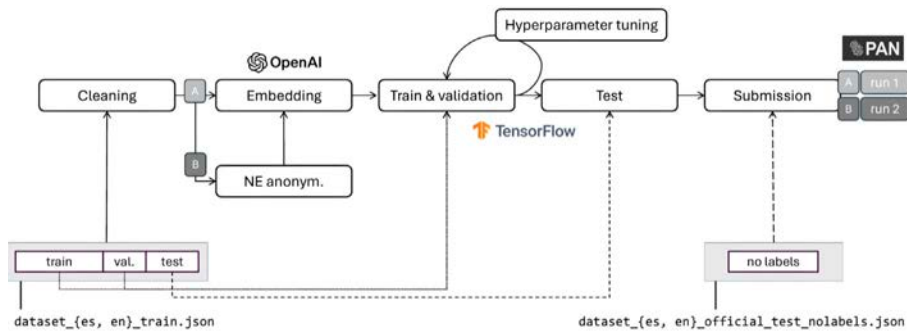


Figura 8.6. Diagrama ilustrativo de la herramienta para detección de mensajes conspirativos en Telegram. Tras la limpieza de datos y la anonimización de entidades, se generan los embeddings, y se entrena y valida el modelo con ajuste de hiperparámetros, para llegar a las pruebas finales antes de la presentación de resultados. Fuente: elaboración propia (Gómez-Romero, González-Silot *et al.*, 2024)

Conclusiones

La IA ha cambiado por completo muchos aspectos de nuestra vida y la desinformación no es ajena a estos cambios: ha transformado cómo se crea desinformación, pero también la forma en la que se lucha contra ella. No obstante, lejos de ser una solución mágica que erradique el problema totalmente, se ha convertido en una herramienta clave para agilizar procesos, detectar patrones y asistir a los verificadores en sus diferentes flujos de trabajo. Si algo hemos aprendido es que la desinformación no es un fenómeno simple, sino un problema en constante evolución que requiere respuestas flexibles y adaptativas. Por eso hemos explorado cómo el enfoque *human-in-the-loop* permite aprovechar la capacidad analítica de la IA sin perder de vista el contexto, la interpretación y el juicio humanos, esenciales en esta lucha. Herramientas como las mostradas demuestran que la clave no es la automatización total, sino la combinación de la velocidad de la IA con el criterio experto de quienes verifican la información.

Aun así, el camino está lejos de ser sencillo. La desinformación se adapta, muta y encuentra nuevas formas de eludir los sistemas de detección. Telegram, por ejemplo, ha demostrado ser un entorno especialmente desafiante, donde la descentralización y la falta de moderación facilitan la propagación de contenido engañoso. Esto nos obliga a desarrollar herramientas que no solo sean precisas, sino también flexibles, capaces de actualizarse constantemente para detectar nuevas narrativas y estrategias de manipulación. Además, los sesgos en los modelos de IA siguen siendo un obstáculo a resolver. Si entrenamos nuestras herramientas con datos limitados o poco representativos, corremos el riesgo

de reforzar prejuicios o ignorar formas sutiles de desinformación. La supervisión humana no es el complemento a estos sistemas, sino el elemento imprescindible para evitar estos problemas.

Y más allá de los retos técnicos, hay un debate mucho más amplio sobre el impacto de la IA en el ecosistema informativo. La lucha contra la desinformación no puede convertirse en un pretexto para imponer restricciones desproporcionadas a la libertad de expresión ni para delegar en algoritmos decisiones que requieren un criterio ético. La transparencia en los modelos, la protección de la privacidad de los usuarios y el acceso abierto a este tipo de herramientas son elementos esenciales para garantizar que estas tecnologías se utilicen de manera responsable. Además, ninguna solución técnica será efectiva si no va acompañada de esfuerzos en alfabetización mediática. La mejor herramienta contra la desinformación seguirá siendo una ciudadanía crítica (que no cínica), atenta y capaz de evaluar la información que consume.

La IA ha llegado para quedarse, pero su papel debe ser el de un asistente, no un juez. Su integración en estrategias híbridas, donde el humano está en el centro, donde la automatización y el conocimiento experto trabajan de la mano, es la clave para construir un ecosistema informativo más fiable. La desinformación no desaparecerá, pero con herramientas bien diseñadas, supervisión humana y una mayor conciencia social, seremos capaces de mitigar su impacto y fortalecer el acceso a una información veraz y contrastada.

Agradecimientos

Este trabajo ha contado con el apoyo del proyecto U-MIND financiado por la Fundación Observatorio Social «la Caixa», el proyecto XAI-DISINFODEMICS (PLEC2021-007681) financiado por el MICIU/AEI/10.13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR y el proyecto Iberifier Plus, cofinanciado por la Comisión Europea en el marco de la convocatoria DIGITAL-2023-DEPLOY-04, Observatorio Europeo de Medios Digitales (EDMO) - Nodos nacionales y multinacionales; referencia: IBERIFIER Plus - 101158511.

Referencias

- Cantón-Correa, J. (2023). *SoundHive: Una propuesta de herramienta de monitorización para verificadores de información*. <https://doi.org/10.5281/zenodo.10015755>.
- Cantón-Correa, J. (2025, enero 10). *Meta dinamita su sistema de verificación externa: ¿barra libre para contenidos peligrosos en redes sociales?* The Conversation. <http://theconversation.com/meta-dinamita-su-sistema-de-verificacion-externa-barra-libre-para-contenidos-peligrosos-en-redes-sociales-246894>.
- Cantón-Correa, J., Sidorenko-Bautista, P. y Alba-Ruiz, R. (2024). Monitorización y detección automática de noticias mediante inteligencia artificial: Una propuesta de herramienta para redes sociales y plataformas. *Espejo de Monografías de Comunicación Social*, 25, Article 25. <https://doi.org/10.52495/c1.emcs.25.p108>.
- Chesney, B. y Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, 107, 1753.
- Gómez-Romero, J., Cantón-Correa, J., Pérez Mercado, R., Prados Abad, F., Molina-Solana, M. y Fajardo, W. (2024). *pytopicgram: A library for data extraction and topic modeling from Telegram channels*. <https://zenodo.org/records/13863008>.
- Gómez-Romero, J., González-Silot, S., Montoro-Montarroso, A., Molina-Solana, M. y Martínez Cámara, E. (2024). Detection of conspiracy-related messages in Telegram with anonymized named entities. *Working Notes of CLEF*. https://downloads.webis.de/pan/publications/papers/gomezromero_2024.pdf.
- Graves, D. (2018). *Understanding the promise and limits of automated fact-checking*.
- Hao, K. (2021, marzo 11). How Facebook got addicted to spreading misinformation. *MIT Technology Review*. <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/>.
- La Barbera, D., Roitero, K. y Mizzaro, S. (2022). A hybrid human-in-the-loop framework for fact checking. *Proceedings of the Sixth Workshop on Natural Language for Artificial Intelligence (NL4AI 2022) co-located with 21th International Conference of the Italian Association for Artificial Intelligence (AI* IA 2022)*.
- Magallón, R. (2019). *Unfaking news: Como combatir la desinformación*. Pirámide.
- Maldita.es. (2024, septiembre 30). *Cómo los canales «alternativos» en Telegram multiplican el alcance de la desinformación y unen a sus seguidores en comunidades*. Maldita.es — Periodismo para que no te la cuelen. <https://maldita.es/nosotros/20240930/canales-telegram-desinformacion-comunidades/>.
- Marín, J. A. N., Aguayo, L. B. y Peralta, R. L. (2023). *Diccionario Posverdad*.
- McIntyre, L. (2018). *Post-truth*. MIT Press. [https://books.google.es/books?hl=es&lr=&id=9-1LDwAAQBAJ&oi=fnd&pg=PP11&dq=%09%E2%80%A2%09McIntyre,+L.+\(2018\).+Post-Truth.+MIT+Press.+&ots=nE9P5uT4lz&sig=bd1l1W5hY-o9pmnnPP65PVgEbKc](https://books.google.es/books?hl=es&lr=&id=9-1LDwAAQBAJ&oi=fnd&pg=PP11&dq=%09%E2%80%A2%09McIntyre,+L.+(2018).+Post-Truth.+MIT+Press.+&ots=nE9P5uT4lz&sig=bd1l1W5hY-o9pmnnPP65PVgEbKc).
- Montoro-Montarroso, A., Cantón-Correa, J., Rosso, P., Chulvi, B., Panizo-Lledot, Á., Huertas-Tato, J., Calvo-Figueras, B., Rementeria, M. J. y Gómez-Romero, J. (2023). Fighting disinformation with artificial intelligence: Fundamentals, advances and challenges. *Profesional de La Información*, 32(3), Article 3. <https://doi.org/10.3145/epi.2023.may.22>.

- Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., Papotti, P., Shaar, S. y Martino, G. D. S. (2021). Automated fact-checking for assisting human fact-checkers. *arXiv preprint arXiv:2103.07769*.
- Newtral. (2022, diciembre 21). ClaimCheck, la herramienta desarrollada por Newtral y ABC Australia para combatir las mentiras repetidas. *Newtral*. <https://www.newtral.es/claimcheck-herramienta-mentiras-repetidas/20221221/>.
- Ufarte-Ruiz, M.-J., Peralta-García, L. y Murcia-Verdú, F. J. (2018). Fact checking: Un nuevo desafío del periodismo. *Profesional de la información*, 27(4), 733-741.
- Vaccari, C. y Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 6(1), 2056305120903408. <https://doi.org/10.1177/2056305120903408>.
- Wardle, C. (2017, marzo 14). *Noticias falsas. Es complicado*. First Draft News ES. <https://es.firstdraftnews.org:443/2017/03/14/noticias-falsas-es-complicado/>.
- Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F. y Choi, Y. (2019). Defending against neural fake news. *Advances in neural information processing systems*32. <https://proceedings.neurips.cc/paper/2019/hash/3e9f0fc9b2f89e043bc6233994dfcf76-Abstract.html>.

Desinformación en contextos tecnológicos y digitales: impacto, marcos legales y estrategias para contrarrestarla

Mónica Niño Romero

Universidad de Huelva
monica.nino@alu.uhu.es

1 | Introducción

En la era digital, la desinformación representa uno de los mayores desafíos para la calidad informativa y la confianza en los medios. Impulsada por algoritmos que priorizan la viralidad sobre la veracidad, su difusión masiva pone en riesgo el funcionamiento de las democracias. Este estudio explora los factores que propician la propagación de desinformación, evaluando marcos legales, tecnológicos y educativos que se han implementado para mitigar su impacto. A través de un análisis multidimensional, se busca ofrecer soluciones para fortalecer la integridad informativa.

En este marco, el presente estudio tiene como objetivo explorar los diversos aspectos de la desinformación en entornos tecnológicos y digitales, analizando los factores que facilitan su difusión, así como las medidas legales y educativas implementadas para contrarrestarla. La investigación aborda, además, la importancia de la alfabetización mediática y la adopción de estrategias tecnológicas que permitan identificar y mitigar la propagación de contenidos falsos. A través de un análisis

exhaustivo de la literatura, el análisis de contenido de proyectos e iniciativas relevantes y el uso de un cuestionario aplicado a periodistas, se busca ofrecer una visión integral del fenómeno y proponer soluciones que contribuyan a fortalecer la integridad informativa.

Este artículo examina, asimismo, el papel de los algoritmos en la difusión de desinformación y el potencial de las tecnologías emergentes, como la inteligencia artificial, para detectar y reducir su impacto. En última instancia, se subraya la necesidad de una colaboración entre gobiernos, plataformas tecnológicas y sociedad civil para crear un marco regulatorio eficaz que equilibre la libertad de expresión con la protección contra la desinformación.

2 | Revisión de la literatura

2.1. Desinformación en el entorno digital

La desinformación, definida como la difusión intencionada de información falsa o engañosa, ha adquirido una nueva dimensión en el entorno digital. La rapidez con la que la información, verídica o no, se propaga a través de las redes sociales y otras plataformas digitales, plantea serios desafíos a la integridad informativa y a la capacidad de las audiencias para discernir la veracidad de los contenidos que consumen (Tandoc *et al.*, 2017). Este fenómeno ha suscitado preocupación entre académicos, profesionales del periodismo y reguladores, quienes buscan comprender no solo los factores que la propician, sino también las consecuencias que tiene sobre la opinión pública y los mecanismos más eficaces para contrarrestarla (Wardle & Derakhshan, 2017).

Tal como aseguran Elías-Zambrano *et al.*:

El contexto político y cultural en el que nuestras sociedades están inmersas ha puesto de manifiesto que la desinformación se difunde en los entornos digitales a velocidades inusitadas e incontrolables. Los bulos, las mentiras y las informaciones sesgadas circulan por nuestra sociedad con independencia de la gravedad de las situaciones que nos toque vivir (2022: 2).

Como indican López-García y Vizoso (2021), el periodismo tiene que mantenerse en constante adaptación a los avances tecnológicos para no quedar rezagado en la lucha contra la desinformación.

2.2. Papel de los algoritmos y plataformas digitales

El análisis de la literatura sugiere que los factores tecnológicos, especialmente los algoritmos utilizados por plataformas como Facebook, Twitter y

Google, juegan un papel crucial en la difusión masiva de desinformación. Estos algoritmos priorizan el contenido que genera mayor interacción, sin tener en cuenta la veracidad del mismo, lo que facilita la viralización de noticias falsas (Vosoughi *et al.*, 2018). Además, la velocidad a la que se propagan estos contenidos supera con creces la capacidad de las plataformas para detectar y eliminar información incorrecta.

En efecto, en un escenario digitalizado y *online*, la rapidez y alcance de las plataformas digitales o redes sociales han amplificado su propagación, dificultando aún más distinguir la verdad; no hay tiempo para verificar y analizar, solo hay prisa, tal y como describen Allcott & Gentzkow (2017). En el contexto tecnológico, las noticias falsas han cobrado una importancia sin precedentes, intensificando la desinformación y debilitando la confianza en el periodismo y en la información veraz.

De esta manera, el fenómeno de la desinformación se vincula a la proliferación de notificaciones falsas y noticias emotivas, por parte de medios amables o empáticos. Bakir & McStay (2018) analizan el hecho de las noticias falsas junto con la expansión de lo que denominan «información emocional» o «medios empáticos». Este tipo de medios ajusta los contenidos y la información de acuerdo con aspectos emocionales, priorizando las respuestas afectivas del público. Encuentran un altavoz en los medios sociales digitales, en las redes sociales.

Igualmente, el espacio digital actual, de naturaleza líquida, es un caldo de cultivo favorecedor de noticias falsas o informaciones sin contrastar, con falta de rigor periodístico. Se ha conformado una sociedad líquida, en contraposición al anterior sistema analógico, sólido y firme, que impulsa este crecimiento de la desinformación en esta sociedad líquida (Aguaded *et al.*, 2022a).

2.3. Marcos regulatorios

Diferentes países han implementado marcos legales para intentar mitigar el impacto de la desinformación en la sociedad. Por ejemplo, Naciones Unidas y la Organización de Estados Americanos (OEA) adoptaron la *Declaración Conjunta sobre Libertad de Expresión y Noticias Falsas* (Fake news), *Desinformación y Propaganda* en 2017, que recalca que la libertad de expresión debe respetarse sin fronteras y que la interferencia en señales de medios solo será legítima bajo decisión judicial imparcial. Además, promueve que los medios adopten sistemas de autorregulación para garantizar la veracidad de las noticias y que ofrezcan cobertura crítica de la desinformación y propaganda, especialmente en momentos clave como elecciones y debates públicos, según Bukovska (2018).

Asimismo, la Unión Europea ha desarrollado un Código de Prácticas contra la Desinformación, que establece directrices para que las

plataformas digitales limiten la difusión de contenido falso (Reglamento (UE) 2022/2065). En el caso de España, se legisla la Ley 13/2022, de 7 de julio, General de Comunicación Audiovisual, que contempla códigos de buenas conductas para la protección de los usuarios respecto de la desinformación.

Sin embargo, la aplicación de estas regulaciones es desigual y enfrenta numerosos desafíos, entre ellos la necesidad de proteger la libertad de expresión.

2.4. Alfabetización mediática y educación

Un aspecto clave en la lucha contra la desinformación es la educación y la alfabetización mediática. Diversos estudios refuerzan esta tesis, tal y como señalan Pérez-Rodríguez y Delgado-Ponce (2012), quienes indican que se debe proveer a la ciudadanía de las facultades para los adecuados procesos de acceso y búsqueda de información en internet, así como el conocimiento sobre la tecnología y los códigos de los mensajes que recibimos.

Por ello, se hace necesario dotar y habilitar a las audiencias en competencias mediáticas y digitales para capacitarles en facultades de acceso a los medios, así como formarse en sentido crítico y con responsabilidad y manejar con actitud y destreza el espacio digital sin debilidades, desde el punto de vista de Aguaded y Romero-Rodríguez (2015).

De igual forma, Benesch (2023) sugiere la necesidad de aplicar enfoques proactivos y flexibles para enfrentar los efectos negativos de la desinformación en línea, como la promoción de la alfabetización mediática, que permite a las personas desarrollar habilidades para analizar la información de manera crítica. Esto contribuiría a identificar mejor la desinformación y los intentos de manipulación.

3 | Metodología

El objetivo principal de este trabajo es analizar la desinformación en contextos tecnológicos y digitales, así como su impacto en la opinión pública, los marcos legales establecidos para combatirla y las medidas metodológicas que se pueden aplicar para su estudio. Los objetivos específicos incluyen:

- Identificar los factores tecnológicos que facilitan la viralización de noticias falsas.
- Explorar el impacto de la desinformación en la formación de la opinión pública.
- Revisar los marcos legales existentes para contrarrestar la desinformación en los medios digitales.

- Examinar los enfoques metodológicos aplicados en el estudio de la desinformación.
- Proponer medidas educativas y tecnológicas que ayuden a defender a las audiencias.

El diseño metodológico de este estudio combina tres enfoques: la revisión de literatura, el análisis de contenido y la aplicación de un cuestionario dirigido a periodistas.

3.1. Revisión de la literatura

En primer lugar, se realizó una exhaustiva revisión de la literatura sobre previos relacionados con la desinformación, centrándose en los aspectos tecnológicos y digitales del fenómeno, así como en las respuestas regulatorias y educativas que han surgido en diferentes contextos. Se ha llevado a cabo la selección de estudios rigurosos y de impacto, con un rango temporal lo más actual posible incluyendo la literatura más relevante. Las fuentes utilizadas incluyen artículos científicos, informes de organismos internacionales, libros especializados y estudios de caso recientes.

3.2. Análisis de contenido

Además, se llevó a cabo un análisis de contenido cualitativo sobre ejemplos recientes de desinformación en medios digitales y redes sociales, identificando patrones en la viralización de noticias falsas, así como las estrategias utilizadas para desacreditarlas. Estas iniciativas se han seleccionado bajo el criterio de su representatividad, independencia, solidez y experiencia en el mundo del periodismo y la comunicación.

Asimismo, en esta esta revisión se consideraron las propuestas de diversas autoridades audiovisuales independientes y medios de comunicación en el ámbito europeo y americano, con públicos objetivos diversos (menores, jóvenes, grupos vulnerables, personas de edad avanzada, etc.). Dichos proyectos e iniciativas son:

- *EduMedia Test*. Entidade Reguladora para a Comunicação Social (ERC), de Portugal (Europa).
- *EduCac*. Consejo Audiovisual de Cataluña (CAC), de España (Europa).
- *Mayores con Wifi*. Consejo Audiovisual de Andalucía (CAA), de España (Europa).
- Alfabetización mediática y formación. Consejo Consultivo de Radio y Televisión (CONCORTV), de Perú (América).

- *Público na Escola. Público*, de Portugal (Europa).
- *Mentes AMI*. (Fundación Atresmedia) de España (Europa).
- *The Learning Network. The New York Times*, de Estados Unidos (América).

3.3. Cuestionario dirigido a periodistas

Finalmente, se diseñó un cuestionario dirigido a periodistas, con el fin de recopilar su percepción sobre el fenómeno de la desinformación. Este cuestionario se ha distribuido a través de plataformas digitales y mensajería instantánea entre profesionales en ejercicio y se ha obtenido una muestra de 114 individuos. Las preguntas se enfocaron en el nivel de alfabetización algorítmica de los encuestados, su percepción sobre el impacto de las tecnologías emergentes en su labor diaria, su disposición para adoptar nuevas herramientas tecnológicas y los desafíos relacionados con su implementación.

4 | Resultados y discusión

4.1. Resultados del análisis de contenido

Los resultados muestran que se han logrado resultados relevantes en las iniciativas analizadas. El proyecto *EduMediaTest* de la ERC ha sido crucial en la promoción de la alfabetización mediática en Portugal. Su objetivo es fomentar competencias críticas, particularmente en la identificación de noticias falsas y el análisis del impacto algorítmico en la difusión de información. Incluye recomendaciones para combatir la desinformación (fig. 9.1.). Esta iniciativa ofrece una base educativa sólida que podría replicarse en otras regiones para mejorar la comprensión de las audiencias sobre los riesgos de la desinformación en entornos digitales. Este programa ha sido llevado a cabo también por el CAC, como un planteamiento a nivel europeo de alfabetización mediática.

El proyecto *EduCAC* es una plataforma educativa que se dirige tanto a estudiantes como a docentes. Desarrolla contenidos sobre la influencia de los algoritmos en la personalización de noticias y su impacto en la creación de burbujas informativas, actuaciones sobre *fake news*, cómo sobrevivir a la desinformación, consejos de verificación antes de compartir contenidos, etc. Esta iniciativa forma parte de los esfuerzos del CAC por aumentar la conciencia crítica sobre el consumo de información digital y los riesgos de la desinformación. En la figura 9.2 se muestran ejemplos de detección de desinformación.

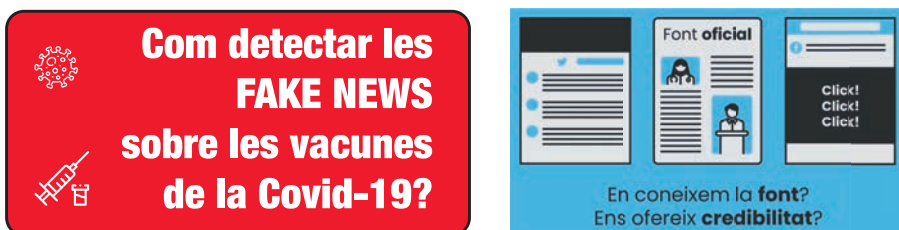
Recomendação 1: O EduMediaTest revela que os alunos possuem uma literacia digital adequada no que concerne às competências técnicas (acesso a conteúdos, gestão de conteúdos, navegação na Internet). Esta conclusão é fortemente indiciada pela dimensão Tecnologia, aquela em que os alunos obtiveram a pontuação geral mais elevada. Esta evidência vem salientar a contradição inerente ao facto de as escolas reduzirem a literacia mediática a uma mera competência digital (essencialmente na sua dimensão instrumental). Por outras palavras, as escolas tendem a oferecer a maior parte de sua formação na única dimensão em que os seus alunos já possuem capacidades. Portanto, torna-se vital continuar a investir na melhoria das aptidões digitais dos alunos através de esquemas educativos que se concentrem na transformação digital, sobretudo sendo esta necessária em contexto de modelos de aprendizagem mista, como a sala de aula invertida (flipped classroom) e o ensino à distância.

Recomendação 2: A dimensão Estética surge em segundo lugar em termos de pontuação. O objetivo desta dimensão é avaliar a capacidade dos alunos em relacionar produções mediáticas com outras obras artísticas, sabendo detetar influências mútuas. Também se destina a avaliar a sensibilidade dos alunos para as vertentes estética e criatividade. É fundamental que os cursos de estética (a importância do desenho, o significado da música, as cores, etc.) sejam valorizados e mantidos para estudantes todas as idades, procedendo-se a um incremento de referências nacionais e interculturais.

Recomendação 3: A dimensão Ideologia posiciona-se em terceiro lugar na pontuação obtida pelos alunos. As questões nela incluídas exploram a forma como as representações dos media estruturam a perceção da realidade, colocando especial atenção no poder da comunicação visual. Esta dimensão envolve questões como: saber distinguir informação fidedigna da informação falsa e explora ainda o fenómeno da desinformação. Por isso, é importante aperfeiçoar o currículo escolar com maior volume de temas relacionados com a conceção ideológica e com a forma como os media operam e causam impacto nos públicos (isto no que respeita às fake news).

Figura 9.1. EduMedia. Fuente: EduMedia Test

Si vols trobar més informació sobre fake news i COVID, la trobaràs en la secció **eduCAC i COVID**, a l'apartat **Famílies** de l'eduCAC.



Com detectar les FAKE NEWS sobre les vacunes de la Covid-19?

Fake news i vacunes de la COVID-19

Com fer front a la desinformació sobre les vacunes de la COVID-19? Les 6 preguntes que ens hauríem de fer.

Font oficial

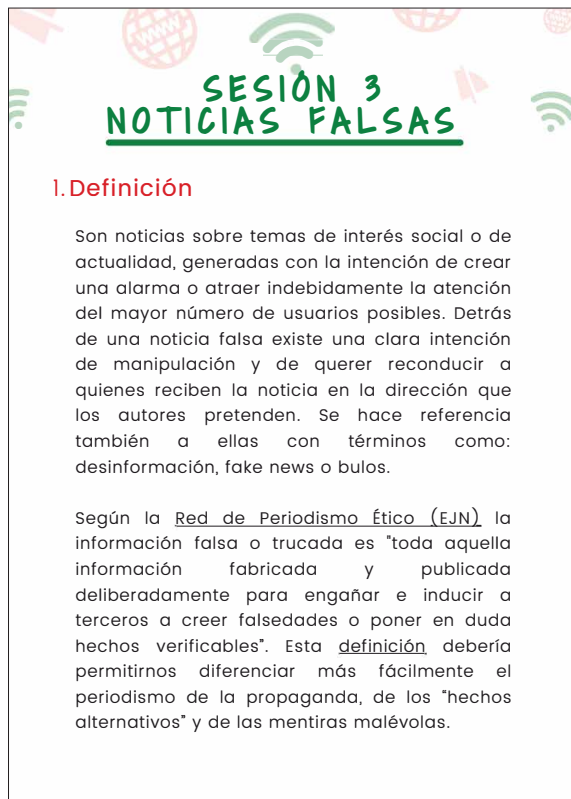
En coneixem la font? Ens ofereix credibilitat?

Verifica abans de compartir

Quins passos podem seguir per verificar qualsevol informació que rebem a través de xarxes de missatgeria?

Figura 9.2. EduCAC. Fuente: Consejo Audiovisual de Cataluña. EduCac <https://www.educac.cat/>

El programa *Mayores con Wifi* del CAA está diseñado para capacitar a personas mayores en la identificación de noticias falsas y en la comprensión de cómo los algoritmos influyen en la información que consumen. Este programa es un ejemplo de la importancia de la alfabetización mediática en grupos vulnerables, y subraya la relevancia de la educación continua para contrarrestar la desinformación, tener en cuenta las noticias falsas, los bulos, entre otros. En la figura 9.3, se presenta un extracto del *Manual didáctico*.



SESIÓN 3 NOTICIAS FALSAS

1. Definición

Son noticias sobre temas de interés social o de actualidad, generadas con la intención de crear una alarma o atraer indebidamente la atención del mayor número de usuarios posibles. Detrás de una noticia falsa existe una clara intención de manipulación y de querer reconducir a quienes reciben la noticia en la dirección que los autores pretenden. Se hace referencia también a ellas con términos como: desinformación, fake news o bulos.

Según la Red de Periodismo Ético (EJN) la información falsa o trucada es "toda aquella información fabricada y publicada deliberadamente para engañar e inducir a terceros a creer falsedades o poner en duda hechos verificables". Esta definición debería permitirnos diferenciar más fácilmente el periodismo de la propaganda, de los "hechos alternativos" y de las mentiras malévolas.

Figura 9.3. Mayores con Wifi. CAA. Fuente: Consejo Audiovisual de Andalucía, <https://consejoaudiovisualdeandalucia.es/2021/10/29/mayores-con-wifi-manual-didactico/>

En el contexto americano, el CONCORTV peruano ha trabajado en diversas iniciativas de alfabetización mediática en Perú, con énfasis en la identificación de desinformación y la promoción de una cultura informativa responsable. Uno de sus programas ha sido *El ABC de la Alfabetización Mediática e Informativa (AMI)* (fig. 9.4). Aunque su enfoque es local, sus proyectos reflejan un compromiso con la transparencia y la formación crítica de las audiencias, contribuyendo a la creación de un ecosistema informativo más saludable, además de propiciar la conectividad digital y la integración en todo el territorio por su diversidad orográfica.

Asimismo, el panorama de los medios de comunicación europeos ha revelado que, en el caso de *Público na Escola* (fig. 9.5), se fomenta la alfabetización mediática entre estudiantes mediante el uso de recursos periodísticos. Este proyecto enseña a los jóvenes a analizar y evaluar críticamente la información, promoviendo un uso consciente de los medios y un rechazo activo a la desinformación.



CONCERTV
Consejo Consultivo de Radio y Televisión

Inicio Nosotros Noticias Investigaciones Eventos Normatividad Boletines

Concertv inicia capacitaciones sobre alfabetización mediática e informacional (AMI)

1 Abril, 2022

EL ABC DE LA ALFABETIZACIÓN MEDIÁTICA E INFORMACIONAL

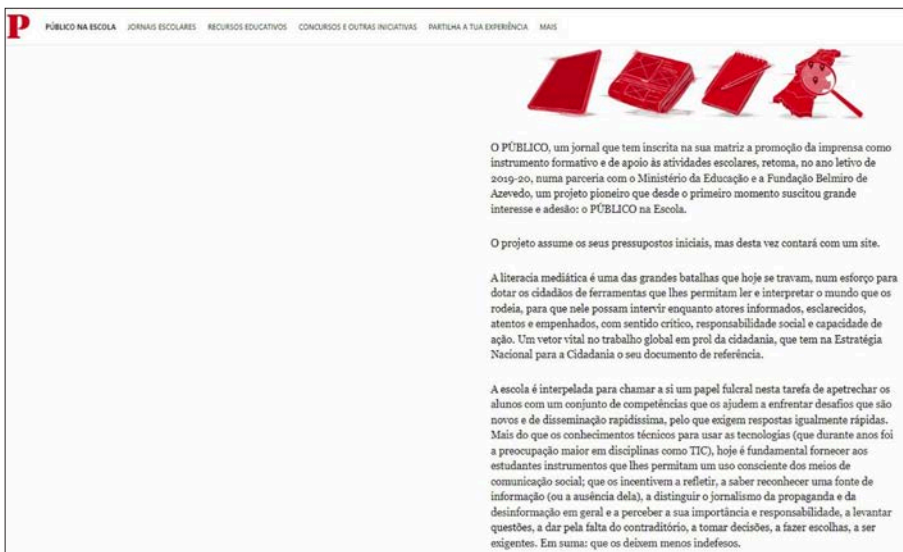
Video thumbnails showing participants in a training session.

Enter keywords

ÚLTIMAS NOTICIAS

- Congreso aprueba modificar Ley de Publicidad Estatal 27 Junio, 2022
- Voces Escolares, experiencia radial que fortalece el protagonismo de adolescentes 25 Junio, 2022
- Canal educativo USMPTV lanza programa en lengua de señas 25 Junio, 2022
- Concertv capacitará a docentes en el análisis crítico del humor televisivo 23 Junio, 2022
- La alfabetización mediática genera empoderamiento 23 Junio, 2022

Figura 9.4. Mayores con Wifi. CONCERTV. Fuente: CONCERTV. <https://www.concertv.gob.pe/>



P PÚBLICO NA ESCOLA JORNAL ESCOLARES RECURSOS EDUCATIVOS CONCURSOS E OUTRAS INICIATIVAS PARTILHA A TUA EXPERIÊNCIA MAIS

O PÚBLICO, um jornal que tem inscrita na sua matriz a promoção da imprensa como instrumento formativo e de apoio às atividades escolares, retoma, no ano letivo de 2019-20, numa parceria com o Ministério da Educação e a Fundação Belmiro de Azevedo, um projeto pioneiro que desde o primeiro momento suscitou grande interesse e adesão: o PÚBLICO na Escola.

O projeto assume os seus pressupostos iniciais, mas desta vez contará com um site.

A literacia mediática é uma das grandes batalhas que hoje se travam, num esforço para dotar os cidadãos de ferramentas que lhes permitam ler e interpretar o mundo que os rodeia, para que nele possam intervir enquanto atores informados, esclarecidos, atentos e empenhados, com sentido crítico, responsabilidade social e capacidade de ação. Um vetor vital no trabalho global em prol da cidadania, que tem na Estratégia Nacional para a Cidadania o seu documento de referência.

A escola é interpelada para chamar a si um papel fulcral nesta tarefa de aptechar os alunos com um conjunto de competências que os ajudem a enfrentar desafios que são novos e de disseminação rapidíssima, pelo que exigem respostas igualmente rápidas. Mais do que os conhecimentos técnicos para usar as tecnologias (que durante anos foi a preocupação maior em disciplinas como TIC), hoje é fundamental fornecer aos estudantes instrumentos que lhes permitam um uso consciente dos meios de comunicação social; que os incentivem a refletir, a saber reconhecer uma fonte de informação (ou a ausência dela), a distinguir o jornalismo da propaganda e da desinformação em geral e a perceber a sua importância e responsabilidade, a levantar questões, a dar pela falta do contraditório, a tomar decisões, a fazer escolhas, a ser exigentes. Em suma: que os deixem menos indefesos.

Figura 9.5. Público na Escola. Fuente: Público. <https://www.publico.pt/publico-na-escola>

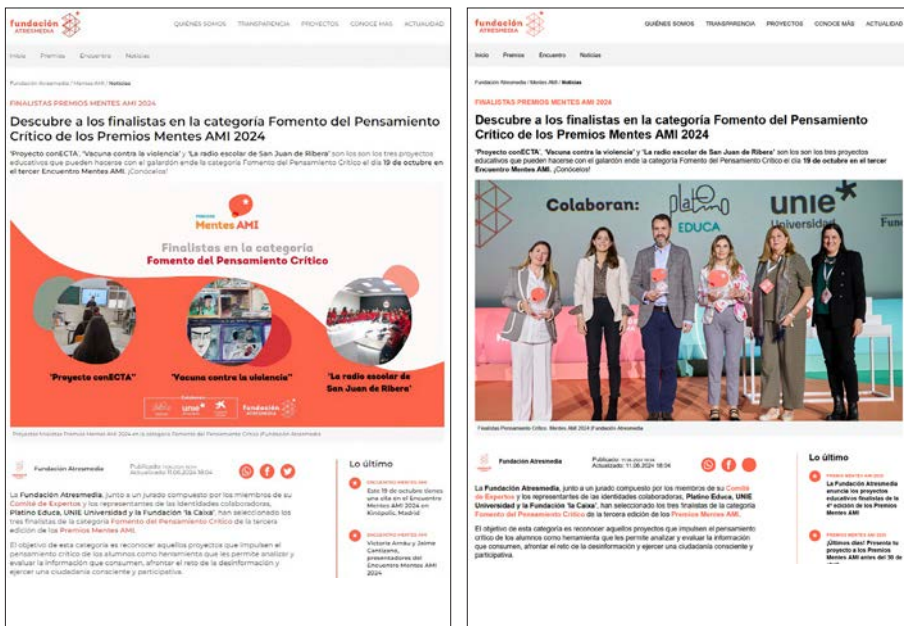


Figura 9.6. Premios Mentes AMI. Verificación y desinformación. Fuente: Fundación Atresmedia. <https://fundacion.atresmedia.com/Mentes-AMI/>



Figura 9.7. The Learning Network. Fuente: The Learning Network. <https://www.nytimes.com/section/learning>

En el caso de *Mentes AMI*, este proyecto fomenta la alfabetización mediática y el pensamiento crítico en torno a la desinformación. Su objetivo es capacitar a los jóvenes para reconocer noticias falsas y entender cómo los algoritmos y las plataformas digitales influyen en la creación y difusión de la información (fig. 9.6).

En el ámbito americano, *The Learning Network* (fig. 9.7) proporciona herramientas educativas a estudiantes y profesores para analizar

críticamente los medios de comunicación. Este proyecto ayuda a los jóvenes a desarrollar habilidades de pensamiento crítico necesarias para detectar desinformación y comprender la importancia del periodismo veraz en un entorno digital en constante cambio.

4.2. Resultados del cuestionario

En relación con la desinformación, los periodistas (fig. 9.8) consideran que la utilización de algoritmos para la verificación de información es algo efectiva (41,6%) y el 25,7% la considera muy efectiva. Por lo tanto, podría contemplarse como una herramienta útil en el espacio de la desinformación y verificación de información.

Para los periodistas consultados, se contempla el uso de herramientas de verificación de información como habilidades esenciales en el paradigma de la alfabetización algorítmica en el periodismo; también —como se ha mencionado anteriormente— en lo relativo a la verificación y desinformación, sobre el uso de algoritmos para verificar la información; además de tener en cuenta la innovación sobre la inteligencia artificial aplicada al periodismo. Acerca de estas habilidades prácticas, los resultados de la muestra (fig. 9.9) destacan las competencias clave necesarias para la alfabetización algorítmica en el periodismo. Entre ellas se encuentran:

- la formación en detección de sesgos y discriminación en algoritmos (50,9%);
- el análisis de datos (47,4%);
- el uso de herramientas para la verificación de información (45,6%);

¿Qué tan efectiva es la utilización de algoritmos para la verificación de información?

125 respuestas

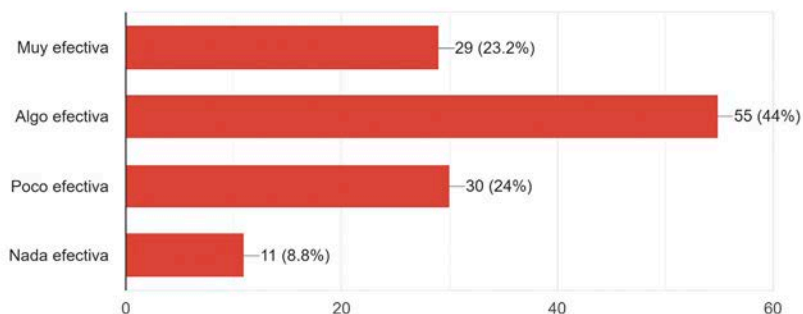


Figura 9.8 Verificación y desinformación. Fuente: elaboración propia. Cuestionario

¿Qué habilidades consideras esenciales para la alfabetización algorítmica en el periodismo?

126 respuestas

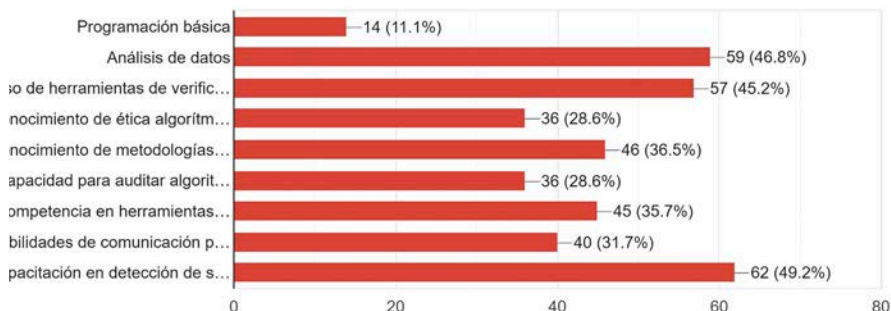


Figura 9.9. Habilidades prácticas. Fuente: elaboración propia. Cuestionario

- d) la capacidad para manejar herramientas de visualización de datos (36,6%);
- e) el conocimiento de metodologías que promuevan la transparencia en el uso de algoritmos (35,1%);
- f) las habilidades comunicativas para explicar al público conceptos relacionados con los algoritmos (31,6%);
- g) la capacidad de auditar algoritmos (28,9%); y
- h) el conocimiento de aspectos éticos relacionados con los algoritmos (27,2%).

En este sentido, el papel de los algoritmos y la inteligencia artificial en la difusión de información es un campo en rápido desarrollo. Si bien estas tecnologías han contribuido a la propagación de noticias falsas, también ofrecen soluciones potenciales. Por ejemplo, se están desarrollando herramientas basadas en inteligencia artificial para detectar automáticamente contenido falso y alertar a los usuarios antes de que lo compartan (Shu *et al.*, 2017). No obstante, estos avances todavía están en sus primeras etapas y presentan limitaciones significativas.

Por otro lado, la asignatura de la capacitación es relevante en el ámbito de los profesionales. La formación profesional en el ámbito del periodismo debe incluir programas que promuevan la capacitación y la sensibilización para convertir la innovación abierta en herramientas prácticas, reduciendo la resistencia frente al cambio tecnológico. En lugar de temer el avance de la tecnología, la mejor estrategia es adquirir nuevas habilidades, ya que la falta de preparación podría comprometer la calidad de los contenidos periodísticos y aumentar la inestabilidad laboral, como señalan Murcia-Verdú y Ufarte-Ruiz (2019). Según Noain-Sánchez (2022), la implementación de nuevas tecnologías más ágiles y versátiles permite reducir tiempos, mejorar la productividad y facilitar tareas

¿Te sientes preparado/a para enfrentar los desafíos que plantean los algoritmos en tu carrera
125 respuestas

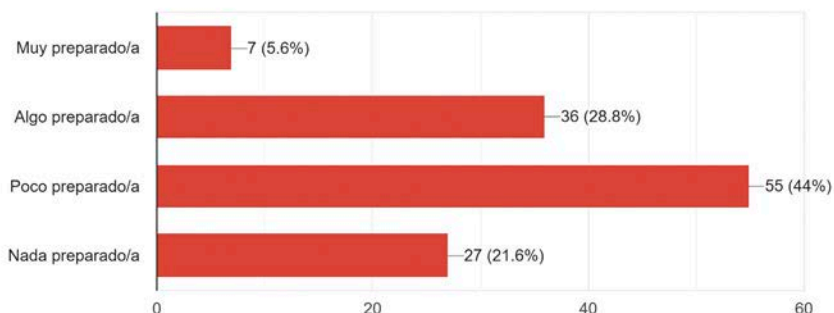


Figura 9.10. Alfabetización en la profesión. Fuente: elaboración propia

repetitivas para los periodistas. La formación y alfabetización tecnológica son esenciales y algunos grandes medios ya han puesto en marcha programas de capacitación a través de sus Laboratorios o «Labs», mediante talleres o plataformas internas, como describen Zaragoza y García-Avilés (2024).

Porque los desafíos son mayúsculos, lo estamos viendo en cuanto a desinformación, tecnologías emergentes y nuevas herramientas. Un ejemplo son los retos que plantean los algoritmos (fig. 9.10): el 43,4% de los profesionales siente que tiene una preparación limitada para enfrentarlos, lo que refleja una falta de confianza en sus competencias actuales y la necesidad de mayor formación en este campo. Un 26,5% se siente moderadamente preparado, indicando que posee un conocimiento básico sobre el impacto de los algoritmos, pero aún necesita mejorar sus habilidades. El 23,9% manifiesta no estar preparado en absoluto, lo que evidencia una preocupación significativa sobre su capacidad de adaptación a las nuevas tecnologías, posiblemente debido a la falta de acceso a recursos formativos. Solo un 6,2% se siente plenamente capacitado, lo que sugiere que este grupo cuenta con una sólida formación y experiencia en el uso de algoritmos, lo que les permite abordar estos desafíos con confianza.

4.3. Espacio digital y desinformación

El análisis empírico confirma las observaciones de Celot (2012) sobre cómo el entorno digital, caracterizado por la sobrecarga informativa y la multiplicidad de pantallas, genera desinformación y pone en riesgo los procesos democráticos. Los resultados del cuestionario dirigido a periodistas muestran que la gran mayoría percibe que la falta de habilidades

críticas en las audiencias facilita la difusión de noticias falsas, lo que concuerda con lo expuesto por Guess *et al.* (2020) sobre la vulnerabilidad de ciertos grupos.

4.4. Influencia en la opinión pública

Los datos del análisis de contenido y la revisión de literatura coinciden en que la desinformación afecta significativamente la opinión pública. Tal como afirman Lewandowsky *et al.* (2017), la exposición continua a información falsa puede crear una «realidad paralela». Esto se refleja en las respuestas de los periodistas, quienes señalaron que la desinformación ha tenido un impacto directo en la percepción de temas políticos y sociales.

4.5. El papel de las redes sociales y la posverdad

El concepto de «posverdad» mencionado por Herrero de la Fuente y Ríos (2021) es respaldado empíricamente: los resultados del cuestionario indican que los periodistas consideran a las redes sociales como las principales fuentes de desinformación. Este hallazgo apoya la idea de que las redes sociales actúan como un «quinto poder» que promueve información no verificada a gran escala. Los periodistas encuestados señalaron que los algoritmos y la falta de regulación en estas plataformas han exacerbado el problema.

4.6. Segmentación y especialización en el periodismo

Redoli (2016) y Tejedor *et al.* (2020) destacaron la segmentación del periodismo digital y la hiperespecialización de contenidos. Los hallazgos empíricos también reflejan esta tendencia, con ejemplos de cómo ciertos medios están creando contenido dirigido a nichos específicos, lo que puede aumentar la polarización y reducir la exposición a perspectivas diversas.

4.7. Marcos regulatorios y efectividad

Los datos sugieren que las regulaciones, aunque importantes, no son suficientes por sí solas. Como mencionan Congreso Futuro (2022) y Lewandowsky *et al.* (2017), las grandes plataformas digitales continúan operando sin un marco regulador global efectivo. Los periodistas encuestados

indicaron que la aplicación de estas regulaciones es desigual y que se necesita una mayor colaboración entre gobiernos y plataformas para reducir la difusión de noticias falsas.

Asimismo, comparamos los marcos regulatorios en tanto que en Europa la implementación de códigos de conducta como el de la Unión Europea ha mostrado algunos avances, pero los periodistas consideran que estas medidas aún no son suficientemente estrictas. En contraste, países como España y Portugal están avanzando con proyectos específicos que promueven la alfabetización mediática, si bien aún no son muy extensivos. Y es cierto que en ámbito americano se ha desarrollado un marco que cubre los principales factores.

4.8. Efectividad de los programas educativos

Los resultados del análisis muestran que los proyectos de alfabetización mediática, como *EduMediaTest* y *Mayores con Wifi*, han sido efectivos en ciertas poblaciones, pero su implementación sigue siendo limitada. Coincide con lo afirmado por Guess *et al.* (2020) sobre la importancia de dotar a las audiencias con herramientas críticas para identificar desinformación.

Sin embargo, los resultados sugieren que la efectividad de estos programas varía. Los periodistas indicaron que los proyectos dirigidos a jóvenes han tenido más éxito, mientras que los dirigidos a adultos mayores y otras audiencias vulnerables necesitan mayor difusión y accesibilidad. Tejedor *et al.* (2020) ya indicaban que se requieren nuevos marcos organizativos para atender esta diversidad de audiencias en la era digital.

Conclusiones

La desinformación en el entorno digital es un problema complejo que requiere respuestas igualmente complejas, considerando todos los factores involucrados. La integración de medidas educativas, regulatorias y tecnológicas es fundamental para mitigar su impacto en la opinión pública y salvaguardar la integridad informativa. En este sentido, es esencial que las audiencias sean educadas para reconocer y combatir la desinformación, mientras que los gobiernos y las plataformas tecnológicas deben trabajar de manera conjunta para desarrollar marcos legales y soluciones tecnológicas efectivas.

Los resultados obtenidos a partir del análisis empírico refuerzan la idea de que la regulación, aunque necesaria, no es suficiente por sí sola. Los marcos regulatorios actuales, tanto a nivel europeo como global, muestran limitaciones en la implementación y en la capacidad para hacer



Figura 9.11. Conclusiones. Estrategias. Fuente: elaboración propia

frente a la rápida difusión de noticias falsas. Por lo tanto, es imprescindible que las estrategias regulatorias sean complementadas con enfoques educativos que promuevan una alfabetización mediática más amplia, tal como se ha evidenciado en los proyectos educativos revisados.

Además, el estudio pone de manifiesto que la educación en alfabetización mediática es particularmente efectiva para fortalecer la resistencia de las audiencias, pero su implementación aún es limitada en muchas regiones. Iniciativas como *EduMediaTest* y *Mayores con Wifi* han demostrado ser herramientas valiosas en la lucha contra la desinformación, pero es crucial expandir estos programas a una mayor diversidad de audiencias y contextos.

Asimismo, es esencial que se continúe investigando y desarrollando metodologías innovadoras para el estudio de la desinformación, con el fin de adaptar las respuestas a las rápidas transformaciones del ecosistema digital. El desarrollo de tecnologías emergentes, como la inteligencia artificial para la detección de noticias falsas, debe ser complementado con una mayor colaboración entre gobiernos, empresas tecnológicas y la sociedad civil.

En conclusión, las estrategias para combatir la desinformación deben seguir un enfoque multidimensional (fig. 9.11). Esto implica una

combinación de medidas regulatorias más estrictas, tecnologías avanzadas de detección de desinformación y, lo más importante, una sólida educación mediática para fortalecer las habilidades críticas de las audiencias. En un entorno digital en constante evolución, la capacidad de adaptarse rápidamente a las nuevas formas de desinformación será clave para salvaguardar la integridad informativa y proteger los procesos democráticos.

Por lo tanto, es vital que las autoridades audiovisuales y los medios sigan colaborando en proyectos que fomenten la educación mediática, con especial énfasis en la alfabetización algorítmica y la detección de desinformación. Finalmente, la implementación de estas estrategias en entornos educativos y la capacitación continua de las audiencias ayudarán a mitigar los efectos de la desinformación en la sociedad.

Referencias

- Allcott, H. y Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211-236. DOI: <https://doi.org/10.1257/jep.31.2.211>.
- Aguaded, I. y Romero-Rodríguez, L. M. (2015). Mediamorfosis y desinformación en la infoesfera: Alfabetización mediática, digital e informacional ante los cambios de hábitos de consumo informativo. *Education in The Knowledge Society (EKS)*: 16, 1, 2015, 44-57. <https://doi.org/10.14201/eks20151614457>.
- Aguaded, I., Pérez-Rodríguez, A. y Delgado-Ponce, Á. (2022). El reto de la alfabetización mediática en la sociedad líquida. *Investigación, desarrollo tecnológico e innovación en educación universitaria*, 17-30. Barcelona, 2022.
- Bakir, V. y McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism*, 6(2), 154-175. <https://doi.org/10.1080/21670811.2017.1345645>.
- Benesch, S. (2023, 2 agosto). *Civil society puts a hand on the wheel: Diverse responses to harmful speech*. Dangerous Speech Project. <https://bit.ly/47j1X1H>.
- Bukovska, B. (2018). *Declaración conjunta sobre Libertad de Expresión y «Noticias Falsas»*.
- Celot, P. (2021). *MEDIA COACH. How to become a media Literacy Coach*. Eavi.eu <https://bit.ly/4dagbna>.
- Congreso Futuro. (2022, 19 enero). *Bruno Patino | No puedo hablar, solo mensajes | Congreso Futuro 2022* [vídeo]. <https://bit.ly/4e1u15y>.
- Del Postigo, D. y Ramírez-Alvarado, M. (2024). Difusión de la ciencia y del conocimiento: líneas de actuación del Consejo Audiovisual de Andalucía en la protección de menores en el entorno digital. *Cuadernos del Audiovisual | CAA*, (11), 11-22. <https://doi.org/10.62269/cavcaa.16>.
- Elías-Zambrano, R., Barrientos, M. y Ramírez-Alvarado, M. M. (2022). Capítulo 6. Medios de comunicación y educación. La necesidad de la Educomunicación en la formación universitaria española. *Espejo de Monografías de Comunicación Social*, (9), 87-101. <https://doi.org/10.52495/c6.emcs.9.p95>.

- Herrero de la Fuente, M. y Ríos, C. (2021). Construcción de un escenario para la posverdad: redes sociales y desinformación. In *Redes sociales, ¿El quinto poder?: una aproximación por ámbitos al fenómeno que ha transformado la comunicación pública y privada* (79-97). Tirant Humanidades.
- Ley 13/2022, de 7 de julio, General de Comunicación Audiovisual, *Boletín Oficial del Estado*, núm. 163, de 8 de julio de 2022, 96114-96220 <https://bit.ly/3Xgx5KK>.
- López-García, X. y Vizoso, A. (2021). Periodismo de alta tecnología: signo de los tiempos digitales del tercer milenio. *Profesional de la información*, 30(3). <https://doi.org/10.3145/epi.2021.may.01>.
- Murcia-Verdú, F. J. y Ufarte-Ruiz, M. J. (2019). Mapa de riesgos del periodismo *hi-tech*. *Hipertext.net*, (18), 47-55. <https://doi.org/10.31009/hipertext.net.2019.i18.05>.
- Noain-Sanchez, A. (2022). Addressing the Impact of Artificial Intelligence on Journalism: the perception of experts, journalists and academics. *Communication & Society-Spain*, 35(3), 105-121. <https://10.15581/003.35.3.105-121>.
- Pérez-Rodríguez, M. A. y Delgado-Ponce, Á. (2012). From digital and audiovisual competence to media competence: Dimensions and indicators. *Comunicar*, 20(39), 25-34. <https://doi.org/10.3916/c39-2012-02-02>.
- Pérez-Tornero, J. M., Tayie, S., Tejedor, S. y Pulido, C. (2018). ¿Cómo afrontar las noticias falseadas mediante la alfabetización periodística? Estado de la cuestión. *Doxa*, 211-235. <https://doi.org/10.31921/doxacom.n26a10>.
- Redoli, D. (2016). Comunicación local para un mundo global. El advenimiento definitivo de la «Glocalización». En *Comunicación local para un mundo global* (13-18). <https://bit.ly/3XoDLqd>.
- Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE. *Diario Oficial de la Unión Europea*. (Reglamento de Servicios Digitales) <https://bit.ly/4dVb3UU>.
- Tandoc, E. C., Lim, Z. W. y Ling, R. (2017). Defining “Fake news”. *Digital Journalism*, 6(2), 137-153. <https://doi.org/10.1080/21670811.2017.1360143>.
- Tejedor, S., Cervi, L., Pulido, C., Perez-Tornero, J. M., Sanjines, D. C., Zhang, W., Tayie, S., Ruiz, J., Vila, P., Sánchez Corral, D., Perez, C., Crespo, E., Mayoral, E., Martin, M. A. y Canovas, P. (2020). *Generación automática de textos periodísticos: La inteligencia artificial aplicada a informativos 2019-2020*. <https://bit.ly/3ryd5aM>.
- Wardle, C. y Derakhshan, H. (2017). *INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making* (Council of Europe report, Ed.).
- Zaragoza, M. T. y García-Avilés, J. A. (2022). Public Service Media laboratories as communities of practice: implementing innovation at BBC News Labs and RTVE Lab. *Journalism Practice*, 1-19. <https://doi.org/10.1080/17512786.2022.2088602>.

Creando una inteligencia artificial capaz de analizar piezas audiovisuales

Eduardo Lleida Solano

Escuela de Ingeniería y Arquitectura (Universidad de Zaragoza)
lleida@unizar.es

Enma Calvo Olloqui

Facultad de CCINF (Universidad Complutense de Madrid)
encalvo@ucm.es

Luis Deltell Escolar

Facultad de CCINF (Universidad Complutense de Madrid)
ldeltell@ucm.es

1 | Introducción

Este proyecto, «Creando una Inteligencia Artificial capaz de analizar piezas audiovisuales», desarrollado entre la Universidad Complutense de Madrid y la Universidad de Zaragoza, busca desarrollar una IA que aúne herramientas para el análisis automatizado de obras audiovisuales. Se pretende que la IA descomponga y extraiga información técnica y narrativa, facilitando el análisis cinematográfico.

El principal objetivo es diseñar una IA que realice tareas de análisis, como la transcripción automática de guiones, la segmentación en planos y el análisis de características clave:

- Clasificar tamaños de planos.
- Detectar movimientos de cámara.
- Identificar la presencia de personajes y clasificarlos según su rol o características visuales.
- Determinar si los planos son de interior o exterior.
- Identificar si se trata de escenas diurnas o nocturnas.

El gran reto es que, aunque existen IA de empresas privadas que realizan algunas de estas tareas, la intención de este trabajo es generar una con código abierto.

La investigación se estructura en dos áreas: cinematografía e informática. La Universidad Complutense define los parámetros cinematográficos que la IA debe reconocer, mientras que la Universidad de Zaragoza desarrolla y entrena el modelo de IA para el análisis audiovisual. Esta colaboración interdisciplinaria es clave, ya que garantiza que las decisiones tecnológicas se alineen con las necesidades analíticas del ámbito cinematográfico.

Este proyecto no ha recibido financiación, pero está integrado dentro de las actividades de ESCINE (Grupo Complutense de Estudios Cinematográficos), el grupo de investigación ViVoLab y el Instituto Universitario de Investigación en Ingeniería de Aragón.

Esta investigación ha sido posible gracias a la ayuda para financiar contratos predoctorales del Estatuto del Personal Investigador en Formación y a la beca de la Residencia de Estudiantes de Madrid, con el apoyo del Ministerio de Ciencia, Innovación y Universidades.

2 | Revisión de la literatura

La IA está llamada a jugar un papel crucial en el estudio y análisis del cine, especialmente en el contexto de la aplicación de métodos científicos y cuantitativos. Aunque se han propuesto una gran variedad de aproximaciones para analizar vídeos de forma automática en las últimas décadas, todavía su aplicación al análisis y estudio de películas es relativamente novedosa. En la literatura se pueden encontrar muchos trabajos e iniciativas relacionados con el análisis de contenidos audiovisuales. Cabe destacar la iniciativa TRECVID (<https://trecvid.nist.gov/>) que, desde 2003, ha sido referencia en el desarrollo de tecnologías para el análisis de contenidos audiovisuales. La llegada en los últimos años de modelos de lenguaje visuales (GPT-4, LLAMA, GEMINI, etc.) ha abierto un abanico de posibilidades, permitiendo combinar la comprensión del lenguaje con el procesamiento de la información visual.

Para el análisis de contenido audiovisual de una película se necesitan tecnologías avanzadas que permitan la transcripción automática de voz a texto, la segmentación de hablantes y reconocimiento de entornos acústicos, la segmentación en planos y el análisis de características clave como la clasificación de tamaños de planos, la detección de movimientos de cámara, la identificación de personajes y su rol, el reconocimiento de actores, la identificación de interiores/exteriores y la determinación de escenas diurnas/nocturnas. Los pocos métodos existentes que intentan esta tarea dependen en gran medida de guiones textuales como entrada, lo que limita en gran medida su aplicabilidad. Recientemente se ha propuesto un sistema, ScreenWriter (Mahon, 2024), para la generación automática de guiones partiendo solo del vídeo y produce resultados que incluyen diálogo, nombres de los personajes, cortes de escena y descripciones visuales.

La transcripción automática de diálogos ha experimentado avances significativos gracias a las redes neuronales profundas, particularmente con modelos basados en arquitecturas Transformer. Modelos abiertos como Whisper de OpenAI (Radford, 2022), aunque no siempre perfectos, demuestran capacidades robustas para diversas calidades de audio y acentos. Algo más problemático es la segmentación por hablantes, aunque la combinación de Whisper con el sistema de segmentación Pyanote (Bain, 2022) da unos resultados aceptables. Como desafíos podemos mencionar el ruido de fondo, la superposición de voces, turnos cortos de diálogo y la jerga específica. Después de la transcripción, las técnicas de procesamiento del lenguaje natural son cruciales para la identificación de personajes por su nombre, el seguimiento de quién y de qué habla y la extracción de información relevante del guion para alimentar otros módulos de análisis.

Los nuevos modelos de lenguaje grandes (LLMs) como los GPT, LLAMA, GEMMA, GEMINI, etc. y posibles variantes especializadas para dominio cinematográfico pueden usarse para la comprensión del contexto y la inferencia de relaciones entre personajes. La correcta atribución del diálogo a los personajes requiere un análisis sofisticado del contexto y la integración con información visual.

El primer paso en análisis de la información visual es la segmentación en planos, que es llamado en el análisis textual tradicionalmente como «deletreo». Las técnicas convencionales se basaban en diferencias de histograma de color, detección de bordes y otras características visuales de bajo nivel entre fotogramas. Las técnicas modernas utilizan redes neuronales convolucionales (CNNs) y redes recurrentes (RNNs) y *transformers*, para aprender patrones complejos de transiciones de planos (Kar, 2024). El plano es la unidad mínima de análisis. Cada plano se caracteriza por su tamaño, movimiento de cámara, personajes, tipo de entorno visual (interior/exterior, diurno/nocturno) entre otros parámetros.

Tradicionalmente, esta caracterización de los planos se realiza con algoritmos de aprendizaje automático ajustados a cada uno de los parámetros a analizar. Las arquitecturas más utilizadas son redes convolucionales, conjuntamente con redes recurrentes que pueden capturar información temporal y contextual para mejorar la precisión de la clasificación (Li, 2023).

En este proyecto se explora el uso de los nuevos grandes modelos de lenguaje multimodales para realizar ese proceso de detección de planos y su caracterización. En esta primera fase se ha realizado una prueba de concepto para estudiar las posibilidades de estos modelos sin un entrenamiento específico (*zero-shot*). A través de un conjunto de instrucciones (*prompt engineering*), se guía al modelo de lenguaje para analizar una secuencia de *frames* de entrada. Dada la limitación en el número de palabras e imágenes de entrada a un modelo de lenguaje, es necesario realizar un proceso preliminar de segmentación de la película. Nuestra propuesta es realizar una segmentación semántica para la que cada imagen se representa por un vector semántico (*embedding*), de forma que imágenes similares ser representan por vectores semánticos próximos en un espacio D-dimensional, siendo D la dimensión del vector. A través del algoritmo de agrupamiento basado en densidad de puntos, HDBSCAN (Campello, 2013), se segmenta la película en partes, cuyas imágenes representan un mismo concepto. Las imágenes representativas de estas serán la entrada al modelo de lenguaje. Para cada segmento el modelo de lenguaje detectará los diversos planos y dará su descripción en términos de los parámetros que se le indiquen.

3 | Metodología

En una primera fase, se investigan las posibilidades de modelos de lenguaje multimodales para segmentar y analizar material audiovisual («deletreo» básico), sin necesidad de una fase de entrenamiento costosa. Todos los modelos utilizados son de acceso libre. El audio se transcribe y cada frase se asigna automáticamente a un hablante, facilitando el análisis de los diálogos. La información visual se segmenta en planos semánticamente coherentes para su posterior análisis, aprovechando los modelos multimodales

En esta etapa inicial, el cortometraje *Di algo* de Luis Deltell (2004) ha sido utilizado para validar la propuesta. Se ha seleccionado este cortometraje para un análisis profundo por su cercanía y por contar con los partes de cámara, la escaleta de montaje y los datos de posproducción del mismo. Su director es uno de los integrantes de este proyecto. Este material audiovisual ha sido fundamental para probar los modelos en condiciones reales y ajustar nuestras métricas. El cortometraje permite testar y entrenar las funcionalidades de la IA, como la segmentación en

planos y la identificación de patrones narrativos, mientras monitoreamos cómo el sistema responde a una narrativa cinematográfica estructurada.

La metodología utilizada para el análisis fragmentario del cortometraje se basa en los enfoques propuestos por David Bordwell y Kristin Thompson (1995), así como en el trabajo de los dos anteriores autores con la colaboración de Janet Staiger (1997). Asimismo, se toma en cuenta las propuestas de análisis textual de González Requena (2012). No se pretende un análisis cualitativo o interpretativo de la película, si no el paso previo metodológico que consiste en un delecto del lenguaje cinematográfico (González Requena, 2012): la identificación y clasificación de los planos, la monitorización de escenas, el desglose de personajes, el análisis de los movimientos de cámara, el uso del sonido y la observación del tiempo y el ritmo narrativo, entre otros aspectos.

En la segunda fase se ajustarán los modelos a la aplicación específica, requiriendo material audiovisual etiquetado. El método incluye aprendizaje supervisado con segmentación manual para entrenar la IA, seguida de validación por expertos para ajustar el modelo y el entrenamiento de la inteligencia y así mejorar el rendimiento de esta.

Para la transcripción de diálogos, se emplean dos sistemas de código abierto: Whisper para convertir voz a texto y Pyannote para segmentar hablantes y asignar frases. Los resultados preliminares en una base de datos de RTVE arrojan un error del 14% en la transcripción y un 18% en la segmentación de hablantes (<https://catedrartve.unizar.es/albayzin-2022results.html>).

Para la representación semántica de las imágenes se utiliza un modelo preentrenado multimodal que asigna a cada imagen un vector semántico o *embedding* D-dimensional, donde D es 768. En este proyecto utilizamos el modelo abierto CLIP (Contrastive Language-Image Pre-Training) desarrollado por OpenAI (Radford, 2021). Su principal habilidad es entender la relación semántica entre texto e imagen, permitiéndole realizar tareas como clasificar imágenes basándose en descripciones textuales (*zero-shot classification*) y buscar imágenes relevantes dada una consulta textual. La representación semántica de la película (secuencia de *embeddings*), se agrupa en planos semánticos utilizando el algoritmo de agrupamiento HDBSCAN. Este algoritmo descubre, de una forma no supervisada, los grupos de imágenes similares que componen la película. La distribución temporal de estos grupos, proceso de diarización, da como resultado los planos de montaje que serán la entrada al modelo de lenguaje multimodal.

En esta primera fase hemos utilizado como modelo de lenguaje multimodal el modelo GPT-4, dado que representa el estado del arte y nos puede dar una idea clara de las posibilidades de esta metodología de análisis de películas. La instrucción dada al modelo de lenguaje es la siguiente:

Esto es una secuencia de imágenes de una escena de un vídeo. Genera una descripción estructurada de la escena de vídeo, dividida en las siguientes secciones:

- **Fragmentación de la toma:** Describe cómo se edita el vídeo. Identifica todas las tomas individuales creadas por cortes y transiciones
- **Tamaño de la toma:** Para cada toma, define el tamaño utilizando la terminología estándar de la cinematografía (por ejemplo, plano general, plano medio, plano americano, plano medio corto, primer plano, primerísimo primer plano).
- **Movimiento de la cámara:** Indica si hay algún movimiento de cámara (paseo, inclinación, seguimiento) o movimiento significativo de personajes dentro del encuadre. Si el tamaño de la toma cambia debido al movimiento, especifica cómo cambia. Si la toma permanece estacionaria en todo momento, anótalo.
- **Punto de vista (POV):** Para cada toma, identifica el POV. ¿Es un punto de vista subjetivo desde la perspectiva de un personaje? ¿O es un punto de vista objetivo, en tercera persona?
- **Localización:** Enumera todas las ubicaciones que aparecen en el vídeo. Indica cuándo aparece cada ubicación y si es en interiores/al aire libre. Anota la hora del día (día o noche). Destaca las ubicaciones clave de la escena, como las casas de los personajes o lugares emblemáticos.
- **Personajes:** Identifica todos los personajes que aparecen en el vídeo.
- **Resumen global:** Escribe un resumen global de la escena de vídeo.

4 | **Discusión y resultados**

Se quiso llevar a cabo un sistema automatizado de análisis audiovisual que combinara tanto información visual como auditiva para descomponer un producto cinematográfico de forma estructurada. La idea era que este proceso se dividiera en varias etapas que interactúan entre sí para lograr una comprensión multimodal de la obra (fig. 10.1).

Se utilizaría el modelo CLIP (Contrastive Language-Image Pretraining) para obtener una representación semántica de las imágenes. Este modelo generaría una representación vectorial de la imagen (*embedding*) con una dimensión de 768, que capturaría las características visuales y semánticas clave de los fotogramas. A partir de estos vectores semánticos, el sistema podría segmentar y agrupar imágenes visualmente similares. A continuación, las representaciones semánticas obtenidas por CLIP serían procesadas mediante un algoritmo de agrupamiento basado en

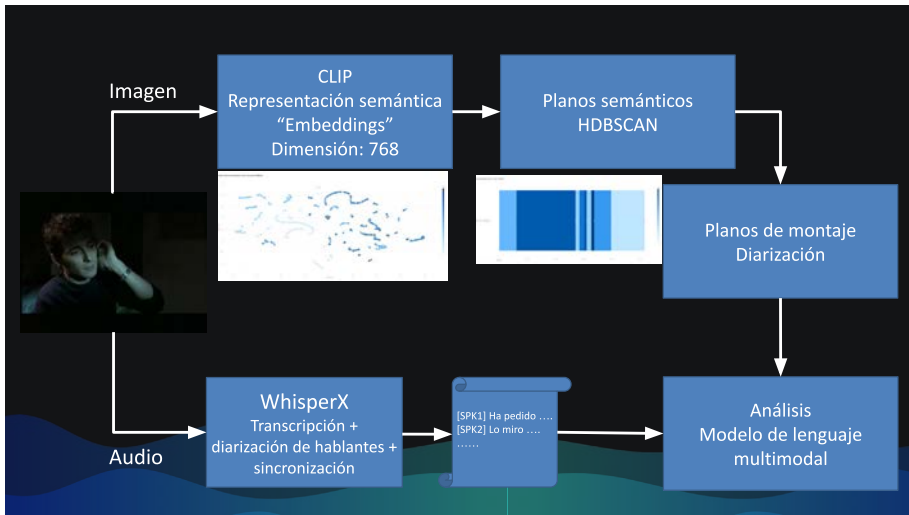


Figura 10.1. Flujo de trabajo para automatizar el análisis audiovisual.

Fuente: elaboración propia

densidad de puntos, HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise). Este algoritmo segmentaría la película en diferentes planos semánticos, agrupando los fotogramas que comparten características visuales y conceptuales similares. Este paso resultaría esencial para la posterior categorización de las secuencias y la identificación de transiciones en la narrativa visual. Paralelamente, el sistema analizaría el componente auditivo utilizando WhisperX, un modelo avanzado para la transcripción automática de diálogos, que también realiza la diferenciación entre los hablantes y sincroniza el audio con la información visual. WhisperX convertiría el audio en texto y lo asociaría con los hablantes correspondientes, asignando etiquetas como [SPK1], [SPK2], etc., para representar a cada interlocutor. El análisis de la transcripción también se incorporaría en el análisis del montaje de la película, determinando cómo se estructurarían los cortes y transiciones entre los planos y las interacciones entre los personajes a lo largo del metraje. Finalmente, tanto la información visual obtenida a partir de CLIP y HDBSCAN como la transcripción de audio de WhisperX, se integrarían en un modelo de lenguaje multimodal. Este modelo sería capaz de realizar un análisis profundo y contextualizado, considerando tanto la dimensión visual como la verbal para ofrecer una descripción integral de la película. A través de esta interacción entre los diferentes tipos de datos, el sistema podría inferir las relaciones entre los personajes, la narrativa y otros aspectos clave de la película.

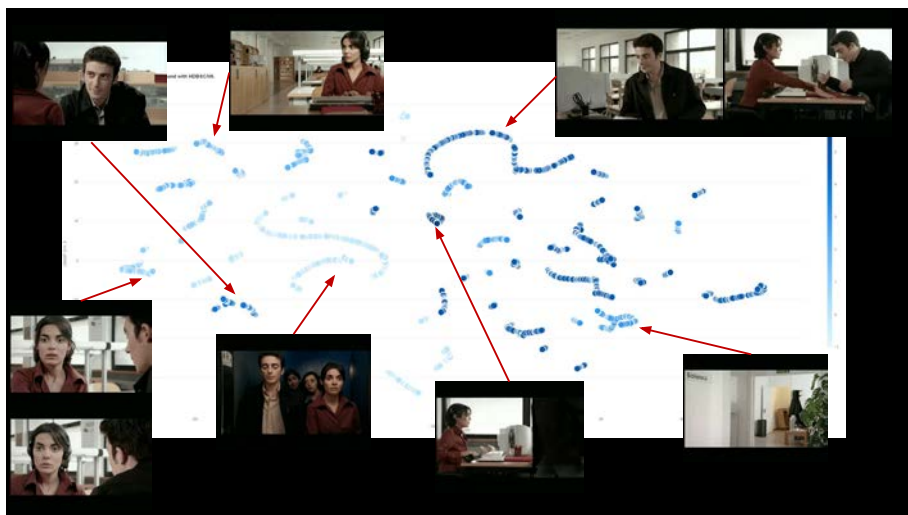


Figura 10.2. Segmentación y agrupamiento de fotogramas. Fuente: elaboración propia

Con ello, se llevaron a cabo las primeras etapas de este sistema. Los resultados de la segmentación obtenida a través de HDBSCAN permitió la segmentación semántica de imágenes y el agrupamiento de planos (fig. 10.2)

Los fotogramas se asociaron con sus representaciones semánticas generadas previamente a través de CLIP, que mapeó cada fotograma en un espacio vectorial de alta dimensión. Las flechas rojas en la imagen señalan cómo las diferentes partes de la película, representadas por las imágenes, se agruparon en el espacio semántico generado por CLIP, reflejando similitudes visuales entre los fotogramas. A partir de las representaciones semánticas de los fotogramas, el algoritmo HDBSCAN organizó los fotogramas similares en planos semánticos. Se visualiza cómo los puntos, que representan los fotogramas, se agruparon y cómo se correlacionaron con las imágenes mostradas en el diagrama. Cada conjunto de puntos en la gráfica tuvo un vínculo con un plano o conjunto de fotogramas con características similares.

A continuación, se concretó el proceso, partiendo de la última escena del cortometraje, elegida por su variedad de planos, personajes y diálogos (fig. 10.3).

Se diferenciaron una serie de planos numerados del 1 al 9, que correspondían a diferentes segmentos del cortometraje. Cada plano fue una unidad básica de análisis visual. Estos planos fueron clasificados en distintas clases semánticas, que se visualizan en el diagrama como columnas de colores. Estas representaban agrupaciones de imágenes similares según su contenido visual. Se agruparon fotogramas visualmente

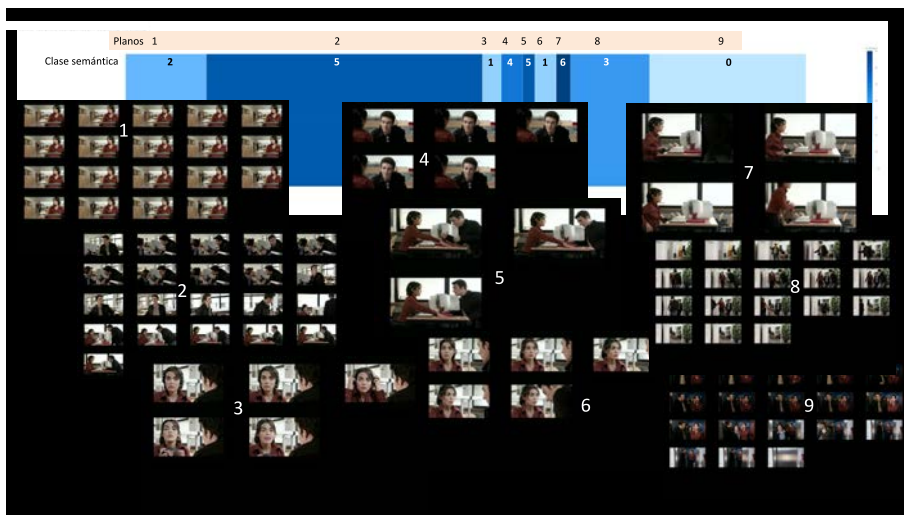


Figura 10.3. Segmentación lineal por planos de la última escena de *Di algo*. Ilustración 3.
Fuente: elaboración propia

similares, lo que permitió identificar patrones recurrentes en la película, como cambios en los planos o escenas similares.

Aunque una evaluación cualitativa inicial mostró buena capacidad para segmentar planos, identificar interiores y exteriores, clasificar tipos de planos y detectar movimientos, no terminó de acertar en otros aspectos. Para hacer una comparativa de aciertos y errores, se utilizaron los datos tomados y desglosados por el equipo humano de este proyecto. Además, se pretende que, en el futuro, se utilicen dichos datos y se amplíen con los obtenidos en otros cortometrajes para que el sistema pueda aprender de ellos y mejorar. Esta fragmentación «deletreada» de las obras audiovisuales es la herramienta básica de trabajo de la metodología de análisis textual. Para este cortometraje de unos 14 minutos de duración, se invirtieron entre 6 y 10 horas de trabajo en hacer esta fragmentación manual. Se elaboró una tabla que segmentó el cortometraje *Di algo* (tab. 10.1). Un ejemplo de ello sería el deletreo que se muestra a continuación de los once primeros planos de la primera secuencia:

Tabla 10.1. «Deletreo», manual del cortometraje *Di algo*

Escena	Plano	Duración del plano	Movimiento ex-terno del plano	Plano según encuadre	Planos según angulación	Plano según punto de vista	Int./Ext.	Día/Noche	Ubicación	Personajes	Texto	Otros
Créditos iniciales		00:00:00:00 • 00:00:25:20									Sí	Créditos iniciales
1	1	00:00:25:21 • 00:00:30:09	No	PD	Picado	Objetivo	Interior	Día	Casa Pablo	Pablo	Sí	Secuencia de montaje
2	2	00:00:30:10-00:00:35:12	No	PD	Picado	Objetivo	Interior	Día	Casa Irene	Irene	Sí(Braille)	
1	3	00:00:35:13 • 00:00:40:16	Panorámica/ Travelling	PD	Picado	Objetivo	Interior	Día	Casa Pablo	Pablo	No	
2	4	00:00:40:17-00:00:53:21	Panorámica/ Travelling	PD	Picado	Objetivo	Interior	Día	Casa Irene	Irene	Sí	
1	5	00:00:53:22 • 00:00:57:05	No	PD	Picado	Objetivo	Interior	Día	Casa Pablo	No	No	
2	6	00:00:57:06- 00:01:03:07	Paneo	PD a PP	Picado a Perfil	Objetivo	Interior	Día	Casa Irene	Irene	No	
1	7	00:01:03:08 • 00:01:12:21	Paneo	PD a PP	Picado a Perfil	Objetivo	Interior	Día	Casa Pablo	Pablo	No	
1	8	00:01:12:22 • 00:01:21:10	No	PD	Picado	Objetivo	Interior	Día	Casa Pablo	Pablo	Sí	
1	9	00:01:21:11 • 00:01:23:23	No	PP	Perfil	Objetivo	Interior	Día	Casa Pablo	Pablo	No	
2	10	00:01:23:24 • 00:01:31:16	No	PMC	Normal	Objetivo	Interior	Día	Casa Irene	Irene	No	
2	11	00:01:31:17 • 00:01:37:09	No	PD	Normal	Objetivo	Interior	Día	Casa Irene	Irene	No	

Fuente: *Di algo* (2004)

Cada fila correspondió a un plano específico, con columnas que incluyeron la escena y número de plano, así como la duración y detalles técnicos como el movimiento de cámara (paneo, *traveling*, etc.), el tipo de encuadre (primer plano, plano medio, etc.), la angulación (cenital, picado, frontal, etc.), y el punto de vista (objetivo o subjetivo). Además, se indicó si la escena era interior o exterior, el momento del día (día o noche), la ubicación (biblioteca, casa, parque, etc.), los personajes presentes, si había diálogo o texto y otros detalles como eventos clave o particularidades en la escena.

A continuación, para hacer esta comparativa entre lo automatizado y lo manual, se utilizaron los resultados obtenidos por el modelo de OpenAi GPT-4. A modo de ejemplo, se muestran a continuación los resultados que se obtuvieron de tres planos de una misma secuencia (figs. 10.4, 10.5 y 10.6).

Estos fueron los resultados obtenidos del primer plano analizado:

— **Fragmentación del plano:**

El video parece ser un solo plano continuo sin cortes ni transiciones.

— **Tamaño del plano:**

El tamaño del plano puede describirse como un plano medio, centrado principalmente en la mujer frente al computador, mientras captura algunos detalles del fondo, como otros escritorios y personas.

— **Movimiento de cámara:**

La cámara permanece estacionaria durante toda la escena. Sin embargo, hay un movimiento significativo de los personajes. Se ve a un hombre caminando desde el fondo hacia el primer plano.

— **Punto de vista (POV):**

La escena se graba desde un punto de vista objetivo en tercera persona. Observa a los personajes sin implicar la perspectiva de un personaje en particular.

— **Ubicación:**

- Ubicación: Un entorno tipo biblioteca u oficina con escritorios y estanterías.
- Tiempo: Interior, durante el día.
- Apariencia: La ubicación permanece consistente a lo largo del video.

— **Personajes:**

- Una mujer sentada en un escritorio usando un computador, con auriculares puestos.
- Un hombre que entra por el fondo y se acerca al primer plano.

— **Resumen global:**

El video muestra una escena ambientada en una biblioteca u oficina. Una mujer está sentada en un escritorio, concentrada



Figura 10.4. Fotogramas de *Di algo 1*. Fuente: elaboración propia



Figura 10.5. Fotogramas de *Di algo 2*. Fuente: *Di algo* (2004)

en su trabajo con auriculares puestos. Mientras tanto, un hombre entra en el plano desde el fondo y se acerca al primer plano, sugiriendo una posible interacción o desarrollo que aún no ha ocurrido. El ambiente es tranquilo y estructurado, típico de un entorno de trabajo o estudio.

Comprobando los resultados automatizados, se constata que la IA acertó al afirmar que: era un único plano, la cámara permanece estacionaria, el punto de vista es en tercera persona, la ubicación y el tiempo.



Figura 10.6. Fotogramas de *Di algo* 8. Fuente: *Di algo* (2004)

Sin embargo, no acierta con el tamaño de plano, dado que es un plano medio corto y el personaje que menciona que se acerca en ningún momento acaba en primer término, además, aparecen esos dos personajes que describe, pero faltaría hacer mención a la bibliotecaria con la que él interactúa. El resumen global es correcto.

Estos fueron los resultados obtenidos del segundo plano analizado:

— **Fragmentación del plano**

- Plano 1: Comienza con ambos personajes en enfoque detrás de un escritorio.
- Plano 2: Se corta a una vista más cercana de los dos personajes en conversación.
- Plano 3: Se corta a otro ángulo similar con ajustes menores en la posición de los personajes.
- Plano 4: Continúa con los dos personajes en el escritorio.
- Plano 5: Se regresa a un plano ligeramente más amplio de los personajes.
- Plano 6: El personaje masculino se aleja del escritorio.
- Plano 7: El personaje masculino se acerca a otro escritorio donde hay un nuevo personaje.
- Plano 8: Comienza una interacción entre el personaje masculino y el nuevo personaje.

— **Tamaño del plano**

- Plano 1: Plano medio de ambos personajes.
- Plano 2: Primer plano medio centrado en la conversación.
- Plano 3-5: Plano medio que mantiene el enfoque en la interacción.

- Plano 6: Plano largo medio mientras el personaje masculino camina hacia otra área.
- Plano 7: Plano medio mientras se acerca al nuevo escritorio.
- Plano 8: Primer plano medio centrado en la interacción con el nuevo personaje.
- **Movimiento de cámara**
 - Plano 1-5: Planos estáticos con ligeros ajustes en los ángulos.
 - Plano 6: La cámara panea ligeramente mientras el personaje masculino se mueve.
 - Plano 7-8: Plano estático con un leve seguimiento para seguir la acción del personaje a través del plano.
- **Punto de vista (POV)**
 - Plano 1-5: Punto de vista objetivo en tercera persona mostrando una interacción mutua.
 - Plano 6: Punto de vista objetivo, capturando el movimiento dentro de la habitación.
 - Plano 7-8: Punto de vista objetivo, centrado en la interacción entre el personaje masculino y el nuevo personaje.
- **Ubicación**
 - Ubicación: Entorno de oficina, interior, durante el día.
 - Tiempo y entorno: Tiempo uniforme a través de todos los planos, mostrando un ambiente de oficina con escritorios de trabajo y computadoras.
- **Personajes**
 - Personaje 1: El personaje masculino, involucrado en conversaciones e interacciones a lo largo de los planos.
 - Personaje 2: Compañera de trabajo femenina en el escritorio, interactuando en la primera serie de planos.
 - Personaje 3: Nuevo personaje femenino encontrado en el segundo escritorio.
- **Resumen global:** El video captura escenas en un entorno de oficina donde un personaje masculino interactúa en conversaciones con dos diferentes compañeras de trabajo. La primera serie de planos establece su interacción con una compañera en un escritorio, centrándose en su diálogo. La segunda parte lo sigue mientras se mueve a otra área de la oficina para conversar con otra compañera, capturando las dinámicas típicas de un entorno laboral. El uso de diferentes tamaños de planos y puntos de vista objetivos mantiene una perspectiva neutral sobre las interacciones que se desarrollan.

En este ejemplo la IA es menos precisa. Era un único plano, pero lo divide en ocho planos diferentes y, con cada uno, realiza el análisis, por lo que este error inicial condiciona los siguientes y los multiplica. Pese a

ello, los resultados conforme el tamaño de planos, el movimiento de cámara, el punto de vista, la ubicación, el tiempo y los personajes son acertados. Es cierto que este es un plano, con movimiento interno y la cámara se mueve siguiendo a uno de los personajes, por lo que todos los elementos van cambiando, razones que pueden justificar su equivoco.

Estos fueron los resultados obtenidos del tercer plano analizado:

— **Fragmentación del plano:**

El video está editado en una secuencia continua con múltiples cortes a medida que los personajes entran y salen del plano, haciendo la transición entre diferentes interacciones de los personajes.

— **Tamaño del plano:**

Los planos parecen mantener un tamaño consistente de «plano medio», capturando a los personajes desde la cintura hacia arriba y el entorno circundante.

— **Movimiento de cámara:**

La cámara permanece estática a lo largo del video, sin movimientos notables de paneo, inclinación o seguimiento. Los cambios ocurren debido al movimiento de los personajes, más que por el movimiento de la cámara.

— **Punto de vista (POV):**

El POV es objetivo y en tercera persona, observando las interacciones entre los personajes sin alinearse con la perspectiva de un personaje en particular.

— **Ubicación:**

- La ubicación es un pasillo interior, posiblemente un espacio de oficina o biblioteca.
- Parece ser de día, indicado por la iluminación brillante y la atmósfera general.
- El entorno incluye mobiliario visible, como un perchero, plantas y un área de asientos con una mesa.

— **Personajes:**

Están presentes varios personajes, incluyendo:

- Un hombre sentado en una mesa leyendo.
- Un personaje masculino entrando en el plano.
- Un personaje femenino entrando junto a otra mujer.

— **Resumen global:** El video muestra una secuencia ambientada en un pasillo interior o área de espera, posiblemente parte de una biblioteca u oficina. Los personajes se mueven por el espacio, participando en breves interacciones. El entorno es casual y profesional, indicado por la decoración y la vestimenta. La escena es capturada desde un punto de vista objetivo, con la cámara permaneciendo fija mientras se desarrolla la acción.

En este último plano analizado, de nuevo, el principal error es ver una fragmentación donde no se produce corte de montaje, al tratarse de un único plano. También se equivoca en la escala de plano dado que se trata de un plano americano. Sin embargo, el movimiento de cámara, el punto de vista, la ubicación, el tiempo y los personajes, son acertados.

Conclusiones

Se presentan las principales conclusiones derivadas de la aplicación de estas tecnologías para la ayuda en el análisis audiovisual, las cuales ofrecen una visión integral de su utilidad y las áreas en las que aún requieren perfeccionamiento.

La IA representa una herramienta valiosa para las personas investigadoras y analistas audiovisuales, ya que facilita un proceso de monitorización y «deletreo» de productos audiovisuales de manera rápida y eficiente. Esto permite reducir significativamente el tiempo dedicado a la monitorización, la contabilización y la extracción de elementos cinematográficos, optimizando así el flujo de trabajo en el análisis audiovisual.

Aunque es cierto que el proceso de descomposición automatizado, o «deletreo automatizado», realizado por la IA aún presenta ciertos errores y fallos que requieren revisión manual por parte del analista. Por tanto, la intervención humana resulta aún necesaria, pero incluso con esta validación manual, el tiempo de análisis se reduce de manera exponencial.

La disponibilidad de modelos de IA de código abierto también ofrece la posibilidad de competir con tecnologías de IA de código cerrado y de pago. Esta accesibilidad democratiza el uso de herramientas avanzadas en el ámbito del análisis audiovisual, permitiendo a los investigadores acceder a recursos de alta calidad sin restricciones económicas.

No obstante, es importante señalar que, en su estado actual (diciembre 2024), la IA no proporciona un análisis cualitativo profundo o relevante. El proceso debe ser entendido como una fase preliminar del análisis, que requiere la intervención y el juicio de la persona investigadora para completar y enriquecer los resultados obtenidos.

Referencias

- Bain, M., Huh, J., Han, T. y Zisserman, A. (2023). WhisperX: Time-accurate speech transcription of long-form audio. INTERSPEECH 2023.
- Bordwell, D. y Thompson, K. (1995). *El arte cinematográfico: Una introducción*. Paidós Comunicación 68 Cine.
- Bordwell, D., Thompson, K. y Staiger, J. (1997). *The classical Hollywood cinema: film style and mode of production to 1960*. Columbia University Press.

- Campello, R. J. G. B., Moulavi, D. y Sander, J. (2013). Density-based clustering based on hierarchical density estimates. En: Pei, J., Tseng, V. S., Cao, L., Motoda, H. y Xu, G. (eds.): *Advances in Knowledge Discovery and Data Mining* (160-172). Springer Berlin Heidelberg.
- González Requena, M. (2012). *Clásico, manierista, postclásico: los modos del relato en el cine de Hollywood*. Castilla Ediciones.
- Kar, T., Kanungo, P., Mohanty, S.N. et al. (2024). Video shot-boundary detection: Issues, challenges, and solutions. *Artificial Intelligence Review*, 57, 104. <https://doi.org/10.1007/s10462-024-10742-1>.
- Li, Y., Lu, T. y Tian, F. (2023). A lightweight weak semantic framework for cinematographic shot classification. *Scientific Reports*, 13, 16089. <https://doi.org/10.1038/s41598-023-43281-w>.
- Mahon, L. y Lapata, M. (2024). ScreenWriter: Automatic screenplay generation and movie summarisation. *arXiv*. <https://arxiv.org/abs/2410.19809>.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., y Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *arXiv*. <https://doi.org/10.48550/arXiv.2103.00020>.
- Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C. y Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. *arXiv*. <https://arxiv.org/abs/2212.04356>.

***Deepfakes* y desinformación organizacional: una revisión de la literatura**

Sónia Gomes Gonçalves

Universidad de Madeira
sonia.goncalves@staff.uma.pt

Pedro Álvaro Pereira Correia

Universidad de Madeira
pacorreia@staff.uma.pt

Las *deepfakes*, una tecnología emergente basada en inteligencia artificial, han captado una atención significativa debido a su potencial para transformar diversos sectores, incluyendo el audiovisual, electoral, periodístico, publicitario, corporativo y el de la investigación científica. Este estudio presenta una revisión de la literatura sobre las *deepfakes* con el objetivo de evidenciar su creciente reconocimiento como una amenaza potencial para múltiples áreas, especialmente en el contexto organizacional.

En un escenario global donde la desinformación se ha convertido en un desafío crítico, las *deepfakes* se destacan como una herramienta capaz de amplificar estos efectos negativos. Los estudios existentes subrayan su impacto en organizaciones, ya sean empresas, marcas, instituciones u organismos, advirtiendo sobre los riesgos relacionados con la reputación, la confianza del público y la seguridad digital.

La revisión se basa en bibliografía emergente en español y portugués, evidenciando que, aunque este tema ha sido escasamente explorado en

comparación con la literatura internacional, ya existe una creciente preocupación sobre las implicaciones de esta tecnología en las organizaciones. Los resultados subrayan la necesidad de más investigaciones profundas y de estrategias efectivas para mitigar los riesgos asociados a las *deepfakes* en diversos contextos.

1 | Introducción

La difusión de información no verificada existe desde siempre, es tan remota como la historia de la civilización (Valero y Oliveira, 2018; Farquhar, 2005; Berkowitz y Schwartz, 2015; Fallis, 2015; Graves y Cherubini, 2016; Allcott y Gentzkow, 2017; Cardoso y Baldi, 2018).

El término «desinformación» tiene poco más de setenta años (Fallis, 2015) y la generalización de su uso ha cristalizado en una polisemia propia del ámbito digital (Magallón-Rosa, 2019). Wardle y Derakhshan (2017) distinguen tres conceptos que describen el desorden de la información: *disinformation* (propagación de información falsa); *misinformation* (intercambio de información errónea) y *malinformation* (difusión de información solo parcialmente real). En la misma línea trabajan Ireton y Posetti (2018: 6), para quienes el concepto de desinformación alude a un intento deliberado y organizado «de confundir o manipular a las personas mediante la entrega de información deshonestas» y Volkoff (2000) quien le atribuye una finalidad asociada a manipulación de la opinión pública con objetivos políticos.

La desinformación enfrenta un nuevo desafío: las *deepfakes*. Han llegado con fuerza y atacan en varios frentes. Vistos como una amenaza para las organizaciones, ya sean empresas, marcas, instituciones u organismos (Westerlund, 2019; Galston, 2020; Cruz *et al.*, 2021; García-Ull, 2021; Brito d'Andréa & Henn, 2021), son usados en guerras de desinformación, pareciendo el equivalente propagandístico de la era de las espadas y los escudos (Chesney & Citron, 2019), porque los contenidos pueden ser peligrosos, de propaganda o de conflicto social (Cerdán Martínez & Padilla Castillo, 2019).

Los *deepfakes* comenzaron a ser usados de forma eficiente en campañas publicitarias, pero pueden representar una amenaza para la política y la seguridad cibernética, aumentando el fraude y el *cyberbullying* (Dasilva *et al.*, 2021).

Atentos a los riesgos, en varios países, inmediatamente se comenzaron a levantar problemas legales sobre su utilización. En los Estados Unidos, varios estados ya han aprobado leyes que regulan las *deepfakes* y otros están preparados para hacerlo. En Europa, el Reglamento de la Inteligencia Artificial ha entrado en vigor en agosto de 2024 y todavía no hay, ni en España ni en Portugal, legislación específica sobre el uso de esta nueva tecnología.

2 | Revisión de la literatura: hipótesis

El presente estudio se desarrolló con base en una investigación exploratoria, a partir de dos revisiones bibliográficas: una sobre el tema genérico *deepfakes* y otra con foco en la desinformación organizacional con recurso a *deepfakes*. Partimos de las siguientes hipótesis:

- Las *deepfakes* son una forma de desinformación;
- Las *cheap fakes* también son *deepfakes*;
- Las *deepfakes* pueden ser usados para propagar desinformación;
- Las *deepfakes* son una amenaza para las organizaciones, ya sean empresas, marcas, instituciones o marcas;
- La difusión de *deepfakes* puede aumentar la desconfianza de los públicos, afectando a las organizaciones;
- Cuando utilizados en el ámbito publicitario, las *deepfakes* pueden (o no) ser benéficos para las organizaciones;
- Esta nueva tecnología preocupa porque representa riesgos a varios niveles: política, seguridad cibernética, fraude y *cyberbullying*;
- Se levantan problemas legales sobre la utilización de *deepfakes*.

3 | Metodología

3.1. Objetivos

Esta investigación se centra en la desinformación para estudiar el impacto de las *deepfakes*, una nueva forma de desinformación. No pretende enumerar los mecanismos de detección de esta nueva tecnología, sino evidenciar que existe conciencia de que las *deepfakes* pueden ser una amenaza para varios sectores, concretamente para las organizaciones, ya sean empresas, marcas, instituciones u organismos. En este sentido, el objetivo general es reunir las evidencias científicas y literarias que permitan evidenciar que se han desarrollado diferentes estudios sobre la temática de estudio.

3.2. Metodología

Este trabajo se desarrolló a través de una investigación exploratoria, partiendo de una breve revisión de la literatura con base en dos revisiones bibliográficas: una sobre el tema genérico *deepfakes* y otra con foco en la desinformación organizacional con recurso a *deepfakes*, partiendo de la base de la responsabilidad social de la información (Sánchez-Gey, Jiménez-Marín, Román-San-Miguel, 2022). Así, primeramente, sin un intervalo de tiempo definido, se partió de una revisión bibliográfica sobre una

temática generalista (las *deepfakes*) con base en literatura nacional e internacional. El criterio de selección de las fuentes y referencias partía del cumplimiento de la pesquisa de ser referencias y autores localizados en revistas, libros o publicaciones encontradas en las bases de datos científica Google Scholar y que enunciaran claramente el término «deepfake». Seguidamente, se analizó la literatura académica emergente sobre desinformación corporativa con recurso a las *deepfakes*. La investigación exploratoria ha permitido seleccionar, durante el mes de febrero de 2022, artículos científicos publicados en español y en portugués, en la plataforma Google Scholar, entre enero de 2021 y febrero del 2022. Los artículos fueron extraídos manualmente y la búsqueda fue restringida a publicaciones durante 14 meses porque se ha sentido necesidad de balizar las pesquisas. Se han utilizado las mismas palabras clave para los dos idiomas y los resultados nos devolvieron un total de 366 artículos. Se ha reducido la muestra excluyendo los libros y los artículos repetidos. Todavía, con recurso al análisis de contenido, esta se ha restringido a estudios que abordan la temática de interés, que está aún poco explorada científicamente. El intuio fue constatar si los artículos referían si las *deepfakes* pueden tener impacto en las empresas, las organizaciones y las marcas. Al final, se recompiló un total de 11 artículos, todos publicados en el año de 2021 (tab. 11.1).

Tabla 11.1. Artículos publicados en la plataforma Google Scholar

Palabras clave	Resultado total	Artículos de interés
« <i>deepfake</i> + empresas»	122	0
« <i>deepfakes</i> + empresas»	128	2
« <i>deep fake</i> + empresas»	60	1
« <i>deep fakes</i> + empresas»	56	4
TOTAL	366	11

Fuente. Elaboración propia

La muestra final usada para el análisis de contenido se constituye por 11 artículos de interés. Siete están escritos en español y cuatro en portugués. Los estudios en español son mayoritariamente de España (tres), siguiéndose Colombia (dos), México (uno) y Argentina (uno). Todas las investigaciones en portugués son de Brasil, con cuatro artículos (tab. 11.2).

Tabla 11.2. Artículos de interés

Idioma	Fecha publicación	Autores	Título	Base de datos	País
ES	Enero 2021	Arencibia, M. G. Cardero, D. M.	<i>Soluciones educativas frente a los dilemas éticos del uso de la tecnología deepfake</i>	Editic - Revista Internacional de Filosofía Teórica y Práctica	Colombia
ES	Enero 2021	Fernández, E. G.	<i>Análisis forense de imágenes y videos</i>	Repositorio INFOTEC	México
ES	Marzo 2021	Gaimari, G.	<i>Inteligencia artificial e impacto en el cibercrimen</i>	Universidad de Belgrano (Tesis - Grado)	Argentina
ES	Junio 2021	García-Ull, F. J.	<i>Deepfakes: el próximo reto en la detección de noticias falsas</i>	Analisi.cat (Universidad Barcelona)	España
ES	Septiembre 2021	Bartolomé, M. C.	<i>Redes sociales, desinformación, cibersoberanía y vigilancia digital: una visión desde la ciberseguridad</i>	RESI: Revista de estudios en seguridad internacional. Universidad de la Rioja	España
ES	Diciembre 2021	Almanza, A. R.	<i>El poder del algoritmo y la vida social</i>	Revista SISTEMAS. Asociación Colombiana de Ingenieros de Sistemas	Colombia
ES	Noviembre 2021	Romero, J. C.	<i>Ciberseguridad: Evolución y tendencias</i>	RESI: Revista de estudios en seguridad internacional. Universidad de la Rioja	España
PT	Marzo 2021	Malheiros, N. D. T.	<i>O Futuro Chegou: Regulamentação Do Uso De Veículos Autônomos No Brasil</i>	UNIALFA - Centro Universitário Alves Faria (Tesis - Maestría)	Brasil

(continúa)

Idioma	Fecha publicación	Autores	Título	Base de datos	País
PT	Mayo 2021	De Brito d'Andréa, C. F. Henn, R	<i>Desinformação, plataformas, pandemia: um panorama e novos desafios de pesquisa</i>	Revista Fronteiras Unisinos - Universidade do Vale do Rio dos Sinos	Brasil
PT	Noviembre 2021	Cafeo, C. G.	<i>Tribunal Superior Eleitoral e o enfrentamento à desinformação nas eleições municipais de 2020</i>	Unesp - Universidade Estadual Paulista (Tesis - Maestría)	Brasil
PT	Octubre 2021	Santaella, L. De Matto Salgado, M.	<i>Deepfake e as consequências sociais da mecanização da desconfiança</i>	TECCOGS. PUC-SP - Pontifícia Universidade Católica de São Paulo	Brasil

Fuente. Elaboración propia

4 | Resultados y discusión

Por un lado, la revisión bibliográfica sobre el tema genérico *deepfakes*, sin un intervalo de tiempo definido, ha permitido diferentes reflexiones sobre la temática, concretamente cuestiones conceptuales; la relación entre las *deepfakes* y la desinformación y el impacto de estos en las organizaciones, pero también en la publicidad y en otros sectores como la política, seguridad cibernética, fraude y *ciberbullying*.

Por otro, el análisis de contenido a la muestra seleccionada, a partir de la bibliografía seleccionada, con foco en la desinformación organizacional con recurso a *deepfakes* y que consistió en una selección bibliográfica de artículos científicos en español y portugués que evidencian el impacto de las *deepfakes* en las empresas, permite concluir que la literatura emergente sobre la temática en estudio es bastante escasa cuando comparada con la publicada a nivel internacional. Todavía hay consciencia sobre los riesgos que el uso de esta nueva tecnología puede causar a nivel corporativo, así como sobre las amenazas que representa a varios niveles. Hay también noción de la importancia de estar en alerta constante, para poder acompañar los avances tecnológicos en esta materia.

Concretamente, se ha verificado que, durante todo el año de 2021, se publicó en la plataforma Google Scholar *más literatura científica sobre la temática en estudio en español que en portugués*. Todavía, el análisis nos permitió verificar que los estudios publicados en español son

mayoritariamente de España, siguiéndose publicaciones aisladas de otros países de América Latina. Por otro lado, todas las publicaciones en portugués son de Brasil.

A continuación se presentan algunas reflexiones sobre las revisiones bibliográficas.

4.1. Deepfakes: cuestiones conceptuales

El término *deepfake* resulta de una combinación de dos términos, *deep* (profundo, pero con origen en *deep learning*) y *fake* (falso). Los *deepfakes* son medios visuales o de audio manipulados o sintéticos que parecen auténticos y que presentan a personas que parecen decir o hacer algo que nunca han dicho o hecho, producidos mediante técnicas de inteligencia artificial, como el aprendizaje automático (*machine learning*) y el aprendizaje profundo (Huijstee *et al.*, 2021).

Un *deepfake* es un vídeo que superpone la cara de una persona en el cuerpo de otra, algo posible gracias a algoritmos gratuitos y fáciles de usar (Cerdán Martínez & Padilla Castillo, 2019).

El Reglamento de Inteligencia Artificial de la UE, que ha entrado en vigor en agosto de 2024, define *deepfake* o ultrasuplantación como «un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objetos, lugares, entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos» [Reglamento (UE) 2024/1689].

El concepto de *deepfake* surge después de la generación de textos automatizados y de la generación de las imágenes falsas. En la primera, con recurso al GPT (transformador generativo preentrenado), que es un nuevo modelo de inteligencia artificial, es posible generar textos que pueden ser usados de forma engañosa como siendo escritos por humanos. En la segunda, gracias a las redes generativas antagónicas (*generative antagonistic network*), se pueden crear rostros de personas que no existen (Giansiracusa, 2021).

En el método usado, el codificador automático extrae las características ocultas de las imágenes faciales y el decodificador se utiliza para reconstruir las imágenes faciales. Para intercambiar caras entre imágenes de origen e imágenes de destino, se necesitan dos conjuntos de decodificadores de codificador en los que cada par se utiliza para entrenar en un conjunto de imágenes y los parámetros del codificador se intercalan entre dos conjuntos de redes (Deshmukh & Wankhade, 2021).

La verdad es que, gracias a la amplia gama de oferta, la técnica usada ha evolucionado y es hoy accesible a todos, desde expertos a principiantes (Deshmukh & Wankhade, 2021).

Es en este contexto que surge el término *cheap fakes*, que es como suele denominarse a las formas de manipulación audiovisual que se basan en *software* barato y accesible o sin recurso a ningún *software* (Paris & Donovan, 2019). Son también *deepfakes*, pero hechos por cualquier persona, con recurso a tecnología ordinaria, o sea, son manipulación audiovisual solo que de forma menos sofisticada (Santaella & Salgado, 2021), pues no requieren de un alto grado de alfabetización digital para su diseño (Gamir-Ríos & Tarullo, 2022).

El primer intento de creación *deepfake* resultó de una aplicación desarrollada en el año 2017 por un usuario de la plataforma Reddit autodenominado «Deepfakes» (Cole, 2017) que usó una técnica de *autoencoder*, es decir, codificador automático (Deshmukh & Wankhade, 2021).

El esquema propuesto por Deshmukh y Wankhade, en que el enfoque permite al codificador común encontrar y aprender la comparación entre dos conjuntos de imágenes faciales, que son relativamente poco desafiantes porque las caras normalmente tienen estructuras similares, como ojos, nariz y lugares para la boca.

El usuario Deepfakes publicó vídeos pornográficos usando los rostros de celebridades como Aubrey Plaza, Gal Gadot, Maisie Williams, Taylor Swift y Scarlett Johansson, consiguiendo popularidad mediante votaciones (Cole, 2017). Aunque omitiendo su verdadera identidad, prestó declaraciones al periódico Vice y explicó que el *software* buscaba imágenes de varios archivos bibliotecarios abiertos y que, para crear las caras de las celebridades, usó la búsqueda de imágenes de Google, fotos de archivo y vídeos de YouTube (*ibidem*).

En la actualidad, existe una amplia variedad de procedimientos de edición de video a través de *deep learning*, por lo que el término *deepfakes* ya no se refiere solo al tipo de intercambio de caras utilizado en las publicaciones originales de Reddit, sobre todo porque desde entonces es cada vez más fácil crear vídeos falsos, manipulados a través de técnicas de inteligencia artificial (Dasilva *et al.*, 2021).

4.2. Deepfakes y desinformación

Desde el punto de vista de la desinformación, las *deepfakes* permiten cumplir algunos objetivos políticos y financieros de los interesados, por lo que se espera que esta nueva tecnología se convierta en el principal proceso de difusión intencional de bulos, sobre todo porque estos son más devastadores que los falsos contenidos de textos, audios o imágenes (Masood *et al.*, 2021).

En los últimos tiempos, la generación *deepfakes* ha avanzado significativamente. Los *deepfakes* con vídeos y audio sintetizados, generados por IA, pueden ser usados para propagar desinformación en todo el

mundo, sobre todo a través de las redes sociales, por lo pueden futuramente representar una grave amenaza (Masood *et al.*, 2021).

Los *deepfakes* se pueden considerar el auge de la desinformación digital y evidencian cambios en el periodismo, porque son la punta del iceberg y dan forma a los desarrollos actuales en el campo de las noticias y los medios. Comprenden fenómenos y desarrollos que incluyen varias formas de desinformación, como noticias falsas, manipulación de los canales de las redes sociales por parte de *trolls* o robots sociales, o incluso la desconfianza del público en la evidencia científica (Huijstee *et al.* 2021).

Por su gran capacidad de desestabilización, se espera un incremento de *deepfakes* en las campañas de desinformación, pues el uso masivo de redes sociales o la enorme cantidad de información disponible acaba creando un efecto denominado infoxicación, que permite multiplicar la efectividad de dichas operaciones (Romero, 2021).

Sobre las *deepfakes*, Maldita.es, un medio de comunicación español dedicado al *fact checking*, que difiere las *deepfakes* de los *cheap fakes* publicó:

Aunque en un futuro podrían convertirse en una amenaza real, en la actualidad hay otras técnicas de desinformación mucho menos complejas, pero igual de efectivas que suponen un peligro mayor. (...) las *deepfakes* podrían convertirse en un problema en el futuro cuando se simplifique y abarate su producción, y cuando se creen herramientas que permitan a cualquier persona hacer uno, pero antes hay que combatir otros tipos de desinformación mucho menos complejas que ya están causando serios problemas en la actualidad. Mientras nos preocupamos con las *deepfakes*, las *cheapfakes* hacen estragos (Maldita.es, 2021).

Pero, en algunas áreas, como en la comunicación social y en la educación, la IA y concretamente las *deepfakes* también se usan para fines no maliciosos. Por ejemplo, la agencia Reuters se ha asociado con la empresa emergente Synthesia con el objetivo de utilizar esta tecnología para interpretar y transmitir informaciones de un juego de fútbol en tiempo real, a través de algoritmos, en un vídeo con un pivote artificial (Casimiro, 2020). Más recientemente, como se previa (Bates, 2020), ha sido noticiado que profesores generados por inteligencia artificial imparten clases en una universidad de Hong Kong, con un avatar de Albert Einstein (EXAME, 2024) y que el uso de *deepfakes* en el periodismo es cada vez más común: en enero de 2024 ha sido lanzado el primer canal de noticias con presentadores generados por IA (Xavier, 2024). También canales de televisión (Pennacchio, 2023; Marcos, 2024) y grandes marcas como IKEA (2024) han recurrido a *deepfakes* para crear presentadores virtuales.

4.3. Impacto de las *deepfakes* en las organizaciones

En el ámbito organizacional, la desinformación gana cada vez más relevancia. Gracias a la tecnología y la inteligencia artificial, los contenidos falsos son usados de forma orquestada para persuadir a los públicos (Gaimari, 2021) y tienen un gran atractivo para los anunciantes, quienes impulsan su difusión como reclamo «porque encaja más fácilmente en los prejuicios de las personas» (Berthon, Treen y Pitt, 2018, p. 4).

La propagación de contenidos falsos es cada vez más peligrosa porque las marcas optan por promover sus productos a través de *influencers* o microcelebridades que explotan la participación dinámica de la cultura web, lo que puede facilitar la información errónea y la desinformación (Baker, 2022). Estos líderes de opinión son percibidos con elevados niveles de fiabilidad e imparcialidad, lo que les permite recomendar productos y servicios en las diferentes plataformas de redes sociales y generar un boca a boca que las marcas aprovechan para su comercialización (De Veirman *et al.*, 2017).

Por otro lado, la difusión de desinformación puede convertirse en un serio riesgo para empresas y marcas (Mut Camacho y Rueda Lozano, 2022) y conllevar un impacto económico negativo (Christov, 2018) que afecte, incluso, a su reputación: según un estudio sobre la *Influencia de las noticias falsas en la opinión pública*, la mayoría de los entrevistados (85,5%) cree que el mayor perjuicio de un contenido falso recae sobre la reputación (Estudio de Comunicación y Servimedia, 2018). Y los contenidos falsos que pueden afectar a la reputación de una empresa obedecen a motivaciones tanto políticas como financieras, emocionales, o ser de carácter interno, como empleados descontentos (Reid, 2017). He aquí la importancia de gestionar relaciones de confianza con los públicos (Xifra, 2020).

Las organizaciones pueden incluso convertirse en víctimas de personas y tecnologías habilitadas con la intención de dañar su reputación con propósitos retorcidos (Peterson, 2020).

Algunos autores encaran la tecnología *deepfake* como una amenaza para las empresas, las organizaciones y las marcas (Westerlund, 2019; Gals-ton, 2020; Cruz *et al.*, 2021; García-Ull, 2021; Brito d'Andréa & Henn, 2021).

Esta tecnología puede ser usada también como una forma de dañar la reputación corporativa, con altas patentes de las empresas y las marcas a supuestamente pronunciarse sobre diversos asuntos (Bécares, 2021).

Es que pueden ser usados sofisticados sistemas de clonación de voz en ataques a las empresas o instituciones gubernamentales (Masood, 2021), así como videos manipulados, para sabotear a los CEO corporativos y sus empresas (Brito d'Andréa & Henn, 2021). Por ejemplo, en marzo de 2019, el director gerente de una compañía de energía británica, creyendo que su jefe estaba hablando con él por teléfono, siguió sus órdenes un viernes por la tarde y transfirió más de 240 mil dólares a una cuenta en Hungría, pero todo no pasaba de una burla y los ladrones

habían utilizado *software* que imitaba la voz y el discurso del ejecutivo máximo de la empresa (Harwell, 2019).

Sobre el riesgo que corren las empresas, ya en el año 2019, noticias financieras y artículos científicos daban cuenta de que estas prácticas podrían aumentar significativamente y de que podrían poner su foco en las empresas, tratando de manipularlas en sus transacciones financieras y decisiones críticas (La Voz del Interior, 2019; Masood, 2021).

En su tercera edición de la revista PTSOC News, dedicada a la seguridad cibernética, la Asociación DNS.PT alerta que las organizaciones están en peligro porque *las deepfakes* pueden ser usadas por criminosos para atacarlas, lo que se podrá traducir en pérdidas financieras (PTSOC News, 2021).

Por otro lado, es importante evidenciar que los medios de comunicación, sobre todo los americanos, así como las empresas de redes sociales y de Internet, como Google, Facebook y Twitter, ya están a tomar las medidas necesarias para lidiar con la propagación de *deepfakes*: los medios se enfocan en capacitar a los periodistas para su detección y las plataformas en línea tienden a financiar proyectos de investigación cuyo objetivo es desarrollar o mejorar herramientas de análisis forense de medios (Vizoso *et al.*, 2021).

Los *deepfakes* son una gran amenaza para la sociedad, para los sistemas políticos y para las empresas porque ejercen presión sobre los periodistas, que luchan por filtrar las noticias reales de las falsas, amenazan la seguridad nacional al difundir propaganda que interfiere en las elecciones y dificultan la confianza de los ciudadanos hacia la información de las autoridades, planteando, todavía, problemas de ciberseguridad para las personas y las organizaciones (Westerlund, 2019).

Es importante señalar que tanto los académicos como la comunidad de IA y el mundo empresarial validan esfuerzos para crear herramientas de detección de *deepfakes* (Wodajo & Atnafu, 2021; Yang, J. *et al.*, 2020; Giansiracusa, 2021; Sun *et al.*, 2021).

No obstante, a medida que la tecnología se desarrolla y se difunde, hay cada vez más técnicas para detectar *deepfakes*. Y en simultáneo esas mismas técnicas son un incentivo para crear *deepfakes* que no sean detectables por estas técnicas (Fallis, 2020).

4.4. *Deepfakes* y publicidad

La publicidad vio rápidamente en las *deepfakes* una oportunidad (Kietzmann *et al.* 2021). Acciones de cariz promocional o publicitario con figuras emblemáticas del pasado comenzaron a ser usadas para promover marcas o productos (Bécares, 2021).

Uno de los ejemplos más emblemáticos es el de una campaña publicitaria lanzada, en enero de 2021, por la marca cervecera Cruzcampo. Esta

fue protagonizada, a través de un *deepfakes*, por Lola Flores, famosa artista española fallecida en 1995, y en cuestión de horas batió récords de difusión en los medios digitales, tornándose viral (Palomo-Domínguez, 2021).

Otro ejemplo es el del comercial de la cadena de supermercados Soriana, que, en mayo del 2021, también con recurso a la tecnología *deepfakes*, «resucitó» a Cantinflas, en un vídeo en el que se ve y oye al artista que falleció el 20 de abril de 1993 (Soriana, 2021).

Es posible que esta nueva tecnología cambie la publicidad (Kietzmann *et al.*, 2020) tal como se la conoce porque las *deepfakes* son más creíbles, accesibles y novedosos que las *fakes news* (Vosoughi *et al.*, 2018).

De verdad, la imaginación es fértil en esta área. En enero de 2020, la empresa internacional Doritos se asoció a Sway para lanzar una aplicación de *deepfakes* que permitía a los usuarios, gracias a un sistema de IA, visualizarse a sí mismos como celebridades bailando en un anuncio con el rapero Lil Nas X y todavía posibilitaba que compartiesen directamente en Instagram, Snapchat, Twitter y TikTok sus movimientos (Kietzmann *et al.*, 2020).

Pero no todo es color de rosa. Es el uso de *deepfakes* en publicidad permite diferentes formas de desinformación. Primero, pueden tomar la forma de información errónea (*misinformation*) convincente y la ficción puede volverse indistinguible de los hechos para un ciudadano común. En segundo lugar, la desinformación (información engañosa creada o distribuida con la intención de causar daño) puede complementarse con materiales falsos para aumentar su potencial engañoso. En tercer lugar, las *deepfakes* se pueden usar en combinación con técnicas políticas de microfocalización, un método publicitario que puede ser muy eficiente porque los productores pueden enviar *deepfakes* personalizados que resuenan fuertemente junto de una audiencia específica (Huijstee *et al.*, 2021).

Las *deepfakes* se difunden fácilmente a través de las redes sociales y de masas y estos medios están tratando febrilmente de gestionar la proliferación de contenido con autenticidad potencialmente engañosa en sus plataformas. Es que las *deepfakes* presentan tanto amenazas como oportunidades para los anunciantes (Kietzmann *et al.*, 2020).

Desarrollos tecnológicos recientes, particularmente aquellos que usan la inteligencia artificial y el aprendizaje automático, desafían la noción contemporánea de publicidad y de contenido publicitario (Campbell *et al.*, 2021; Li, 2019), por lo que es importante pensar en posibles implicaciones más amplias del uso de *deepfakes* antes de apresurarse a integrar la nueva y atractiva tecnología (Kietzmann *et al.*, 2020).

Y la producción y distribución de material publicitario dejó de depender apenas del esfuerzo humano y de herramientas analógicas. Pasó a poder servirse de las innovaciones tecnológicas, que ofrecen una industria publicitaria digital y herramientas automáticas que permiten a los anunciantes automatizar muchos procesos publicitarios y producir

«anuncios sintéticos» o anuncios que comprenden contenido basado en la producción artificial y automática, así como en la modificación de datos (Campbell *et al.*, 2021).

Es que la publicidad sintética se refiere a anuncios que se generan o editan mediante la producción y modificación artificial y automática de datos. Por lo general, a través de algoritmos de inteligencia artificial, que permiten falsificaciones profundas mediante las GAN, creándose automáticamente contenidos que representan una versión artificial y falsa de una realidad muy convincente (*ibidem*).

4.5. Otros riesgos: política, seguridad cibernética, fraude y cyberbullying

El fenómeno de las *deepfakes* se ha convertido en una preocupación para algunos gobiernos porque representa una amenaza no solo para la política, pero también por la seguridad cibernética, aumentando el fraude y el *cyberbullying* (Dasilva *et al.*, 2021).

Un estudio reciente de la internacional Sensity evidencia que las *deepfakes* de vídeo no consensuados y dañinos creados por expertos se duplican aproximadamente cada seis meses. Y hasta diciembre de 2020 fueron detectados más de 85 mil (Sensity, 2021).

Los delitos cibernéticos con recurso a *deepfakes* preocupan bastante:

Las redes generativas antagónicas extendidas al *deep learning* han demostrado su extraordinaria capacidad en los campos de la imagen, el audio y el habla. Pero si la tecnología avanzada nos beneficia, también representa una amenaza para nosotros cuando se usa en delitos cibernéticos (Yang, J. *et al.*, 2020, p. 1).

Las *deepfakes*, además de haber despertado la preocupación de todos los organismos involucrados por su posible uso intencional, también representan una amenaza inminente porque pueden ser utilizados con fines dañinos como el robo de identidad, el *phishing* (técnicas para obtener información confidencial e instalar programas maliciosos) y la estafa (Wodajo & Atnafu, 2021).

Esto ocurre porque las redes generativas antagónicas y las tecnologías de aprendizaje profundo representan una gran amenaza para la seguridad pública e, infelizmente, los métodos tradicionales de detección de falsificación y manipulación no son suficientes para detectar imágenes o vídeos trabajados a través de esta tecnología (Sun *et al.*, 2021). Y está comprobado que los avances en esta área se han empleado para crear *software* que puede causar amenazas a la privacidad, la democracia y la seguridad nacional (Nguyen *et al.*, 2021).

Altas autoridades policiales también se muestran preocupadas y como tal atentas a los peligros que las *deepfakes* representan. Una información del FBI, publicada en marzo de 2021, dejaba alertas sobre la posibilidad de actores maliciosos que utilizan, para operaciones cibernéticas, técnicas de *phishing* e ingeniería social. Esta última se refiere a las técnicas de manipulación psicológica que pretenden conseguir que los usuarios revelen información confidencial o realicen cualquier tipo de acción que pueda beneficiar al ciberdelincuente (INCIDE, 2019):

Los actores extranjeros están utilizando actualmente contenido sintético en sus campañas para influir y el FBI predice que será cada vez más utilizado por actores cibernéticos y delincuentes extranjeros para el *spear-phishing* y la ingeniería social en una evolución del comercio cibernético operativo (FBI, 2021).

También un informe de la Europol sobre la «Evaluación de amenazas de delitos graves y organizados», publicado en abril del mismo año, espera un aumento del uso de *deepfakes* en el crimen cibernético:

Las autoridades policiales tienen poderes limitados para contener la manipulación de información, que puede tomar la forma de intentos de distorsionar el discurso político, manipular las elecciones, erosionar los principios democráticos, sembrar desconfianza en las instituciones, intensificar las divisiones sociales, fomentar la inseguridad y difundir la discriminación y la xenofobia (...). El uso delictivo de la IA, incluida la explotación de *deepfakes*, se espera que aumente en el futuro. La incorporación de la IA a las técnicas existentes puede ampliar el alcance y escalar los ciberataques (Europol, 2021).

4.6. Problemas legales sobre la utilización de *deepfakes*

Como las *deepfakes* fueron usados inicialmente para vídeos falsos de cariz pornográfico y luego para campañas políticas esencialmente difamatorias, inmediatamente se levantaron varios problemas legales sobre su utilización (Kugler & Pace, 2021).

En los Estados Unidos, varios estados ya han aprobado leyes que regulan las *deepfakes* y otros están preparados para hacerlo, por lo que hay una corriente que defiende que las prohibiciones sobre vídeos pornográficos que resulten de *deepfakes* deben recibir el mismo tratamiento, bajo la Primera Enmienda que las prohibiciones de la pornografía no consensual, en lugar de ser tratados bajo la menos protectora ley de la difamación (*ibidem*).

Los mecanismos legales existentes son insuficientes para abordar la amenaza porque la Sección 230 de la Ley de Decencia en la Comunicación

protege a las empresas de redes sociales de la responsabilidad por las *deepfakes* difundidos en sus plataformas. Incluso las enmiendas propuestas a la Sección 230 no abordan adecuadamente la amenaza de las *deepfakes* (...). Las agencias reguladoras deberían promulgar nuevas reglas y enmendar las existentes para responsabilizar a las empresas de redes sociales por la circulación de *deepfakes*. La amenaza de responsabilidad disuadirá a las empresas de redes sociales de permitir que los videos se propaguen sin control en sus plataformas y las incentivará a desarrollar nueva tecnología para su pronta detección y eliminación (O'Donnell, 2021, p. 1).

Por su parte, la Unión Europea ha publicado la Ley de Inteligencia Artificial, designada de Reglamento de la Ley Artificial, que ha entrado en vigor el 1 de agosto de 2024, en el que se establecen normas armonizadas en materia de inteligencia artificial, obligando a transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA:

Los responsables del despliegue de un sistema de IA que genere o manipule imágenes o contenidos de audio o vídeo que constituyan una ultraplantación harán público que estos contenidos o imágenes han sido generados o manipulados de manera artificial. Esta obligación no se aplicará cuando la ley autorice su uso para detectar, prevenir, investigar o enjuiciar delitos. Cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, las obligaciones de transparencia establecidas en el presente apartado se limitarán a la obligación de hacer pública la existencia de dicho contenido generado o manipulado artificialmente de una manera adecuada que no dificulte la exhibición o el disfrute de la obra (Artículo 50, Cap. IV) [Reglamento (UE) 2024/1689].

Los *deepfakes* no están legislados explícitamente ni en España ni en Portugal. Aun así, importa referir que, en España, un partido político ha presentado una Proposición de Ley Orgánica de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial que ha sido tomada en consideración en octubre de 2023 (Proposición de Ley Orgánica 122/000011, BOCG, 2023).

Es importante destacar que la Ley de Servicios Digitales, de octubre de 2023, obliga a que un elemento de información, ya se trate de imagen, audio o vídeo generado o manipulado, «que se asemeja notablemente a personas, objetos, lugares u otras entidades o sucesos existentes y que puede inducir erróneamente a una persona a pensar que son auténticos o verídicos» [Reglamento (UE) 2022/2065 de 19 de octubre] debe ser publicitado con indicaciones destacadas, proporcionando una funcionalidad fácil de utilizar que permita a los destinatarios del servicio señalar dicha información (*ibidem*).

En España, por sus fines, la penalización podría ser por delito contra el derecho a la propia imagen, injuria o delito de odio (Cerdán Martínez & Padilla Castillo, 2019). De esta forma, según en el artículo 18 de la Constitución Española y tipificado en la Ley Orgánica 1/1982, de 5 de mayo, en lo que se refiere a la protección civil del honor, de la intimidad y de la propia imagen, el artículo séptimo establece en sus números 3, 5, 6 y 7 las siguientes intromisiones ilegítimas:

- (3) «La divulgación de hechos relativos a la vida privada de una persona o familia que afecten a su reputación y buen nombre, así como la revelación o publicación del contenido de cartas, memorias u otros escritos personales de carácter íntimo (...)».
- (5) «La captación, reproducción o publicación por fotografía, filme, o cualquier otro procedimiento, de la imagen de una persona en lugares o momentos de su vida privada o fuera de ellos, salvo los casos previstos en el artículo octavo, dos».
- (6) «La utilización del nombre, de la voz o de la imagen de una persona para fines publicitarios, comerciales o de naturaleza análoga».
- (7) «La imputación de hechos o la manifestación de juicios de valor a través de acciones o expresiones que de cualquier modo lesionen la dignidad de otra persona, menoscabando su fama o atentando contra su propia estimación (Ley Orgánica 1/1982, de 5 de mayo)».

En Portugal, en mayo de 2021, se aprobó la Ley n.º 27/2021, con la Carta Portuguesa de Derechos Humanos en la Era Digital. El artículo sexto establece el derecho a la protección contra la desinformación:

1. «El Estado vela por el cumplimiento en Portugal del Plan Europeo de Acción contra la Desinformación, con el fin de proteger a la sociedad frente a las personas singulares o colectivas, de jure o de facto, que produzcan, reproduzcan o difundan una narrativa considerada desinformación, de conformidad con el párrafo siguiente».
2. «Se considera desinformación cualquier narrativa que se demuestre que es falsa o engañosa, creada, presentada y difundida para obtener una ventaja económica o para engañar deliberadamente al público, y que pueda causar un daño público, es decir, una amenaza a los procesos políticos democráticos, a los procesos de elaboración de políticas públicas y a bienes públicos».
3. «A los efectos del párrafo anterior, se considera, entre otras cosas, información demostrablemente falsa o engañosa la utilización de textos o vídeos manipulados o fabricados, así como las prácticas para inundar buzones de correo electrónico y el uso de redes de seguidores ficticios».

4. «Los meros errores en la comunicación de información, así como las sátiras o parodias, no están amparados por lo dispuesto en este artículo».
5. «Toda persona tiene derecho a presentar y ver evaluadas por la Autoridad Reguladora de los Medios de Comunicación denuncias contra las entidades que realicen los actos previstos en este artículo, aplicándose los medios de acción a que se refiere el artículo 21 y las disposiciones de la Ley n.º 53/2005, de 8 de noviembre, sobre los procedimientos de denuncia y deliberación y el régimen sancionador».
6. «El Estado apoya la creación de estructuras para la verificación de hechos por medios debidamente registrados y promueve la atribución de sellos de calidad por parte de entidades confiables y dotadas de la condición de utilidad pública (Diário da República, 17 de mayo de 2021)».

Como las *deepfakes*, además de una amenaza social, pueden representar un peligro para las empresas, las organizaciones y las marcas, urge encontrar formas de combate a varios niveles: detección, regulación, alfabetización mediática y educomunicación, entre otros.

Conclusiones

Se considera que los objetivos planteados en esta investigación se han cumplido plenamente. La revisión de la literatura ha permitido explorar conceptos clave relacionados con las *deepfakes*, destacándolos como una nueva y poderosa forma de desinformación que plantea riesgos significativos. En los últimos años, áreas como las ciencias sociales y los sistemas informáticos han desarrollado investigaciones valiosas sobre este fenómeno (Gamir-Ríos y Tarullo, 2022). Sin embargo, en lo que respecta a la literatura en español y portugués sobre desinformación organizacional mediante el uso de *deepfakes*, el interés de los investigadores y del mundo empresarial es evidente, pero los resultados son aún menos alentadores de lo esperado, considerando la amplia cobertura mediática del tema.

La revisión de literatura ha permitido evidenciar la urgencia de combatir las *deepfakes*, videos manipulados difundidos a través de las redes sociales, que se usan de forma masiva, creándose un efecto denominado de infoxicación (Romero, 2021). Por su nivel de peligrosidad para falsear la realidad, las *deepfakes* son vistos como una preocupación por el daño que provocan (Arencibia y Cardero, 2021), pues pueden ser usados como una poderosa forma de ataque (Gaimari, 2021). Es que pueden representar, además de una amenaza social, un peligro para las organizaciones, ya sean empresas, marcas, instituciones u organismos.

Como los contenidos falsos son actualmente el uso más conocido de la IA como vector de ataque en el cibercrimen, en el futuro, se van a necesitar nuevas tecnologías de detección para mitigar el riesgo de campañas de desinformación y extorsión (Gaimari, 2021). Solo de esta forma y con un esfuerzo conjunto de los ciudadanos, se podrá mantener la confianza en las organizaciones (Bartolomé, 2021; Cafeo, 2021).

Finalmente, este trabajo busca contribuir al conocimiento sobre las *deepfakes* y servir como incentivo para nuevas investigaciones que exploren en profundidad sus implicaciones organizacionales, legales y sociales.

Referencias

- Allcott, H. y Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211-236. DOI:10.1257/jep.31.2.211.
- Almanza, A. R. (2021). El poder del algoritmo y la vida social. *Sistemas*, (161), 24-47. <https://sistemas.acis.org.co/index.php/sistemas/article/view/166>.
- Arencibia, M. G. y Cardero, D. M. (2021). Soluciones educativas frente a los dilemas éticos del uso de la tecnología *deep fake*. *Revista Internacional De Filosofía Teórica Y Práctica*, 1(1), 99-126. <http://riftp.editic.net/index.php/riftp/article/view/22>.
- Bakir, V. y McStay, A. (2017). Fake news and The Economy of Emotions: Problems, causes, solutions. *Digital Journalism*, 6(2), 154-175. https://research.bangor.ac.uk/portal/files/19296816/2017_Fake_news.pdf.
- Baker, S. A. (2022). Alt. Health Influencers: How wellness culture and web culture have been weaponised to promote conspiracy theories and far-right extremism during the COVID-19 pandemic. *European Journal of Cultural Studies*, 25(1), 3-24. DOI: 10.1177/13675494211062623.
- Bartolomé, M. C. (2021). Redes sociales, desinformación, cibersoberanía y vigilancia digital: una visión desde la ciberseguridad. *Revista de Estudios en Seguridad Internacional*, 7(2), 167-185. <https://dialnet.unirioja.es/servlet/articulo?codigo=8306043>.
- Bates, T., Cobo, C., Mariño, O. y Wheeler, S. (2020). Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education*, 17(1), 42. DOI: 10.1186/s41239-020-00218-x.
- Bécares, B. (2021). La historia de cómo un deepfake consiguió dañar la reputación de una empresa de refrescos en España: así son los nuevos ciberdelitos. *Genbeta*. <https://www.genbeta.com/actualidad/historia-como-deepfake-consiguio-danar-reputacion-empresa-refrescos-espana-asi-nuevos-ciberdelitos>.
- Berkowitz, D. y Schwartz, D. A. (2015). Miley, CNN and the Onion: When fake news becomes realer than real. *Journalism Practice*, 10(1), 1-17. DOI: 10.1080/17512786.2015.1006933.
- Berthon, P., Treen, E. y Pitt, L. (2018). How Truthiness, Fake News and Post-Fact Endanger Brands and What to Do About It. *NIM Marketing Intelligence Review, Sciendo*, vol. 10(1), mayo, 18-23. DOI: 10.2478/gfkmir-2018-0003.

- Brito d'Andréa de, C. F. y Henn, R. (2021). Desinformação, plataformas, pandemia: um panorama e novos desafios de pesquisa. *Fronteiras-estudos midiáticos*, 23(2). <http://www.revistas.unisinos.br/index.php/fronteiras/article/view/23786>.
- Cafeo, C. G. (2021). *Tribunal Superior Eleitoral e o enfrentamento à desinformação nas eleições municipais de 2020*. Dissertação de Mestrado. Unesp - Universidade Estadual Paulista. Brasil. <https://repositorio.unesp.br/handle/11449/215521>.
- Campbell, C., Plangger, K., Sands, S. y Kietzmann, J. (2021) Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising. *Journal of Advertising*. DOI: 10.1080/00913367.2021.1909515.
- Cardoso, G., Baldi, V., Pais, P. C., Paisana, M., Quintanilha, T. L. y Couraceiro, P. (2018). As Fake News numa sociedade pós-verdade: Contextualização, potenciais soluções e análise. *Relatórios OberCom*. Junho 2018. OberCom. Lisboa. DOI: 10.13140/RG.2.2.34603.75044.
- Casimiro, J. T. (2020, 13 de fevereiro). 'Deepfakes'. Protótipo recorre à inteligência artificial para marcar nova era da informação digital. *O Jornal Económico*. <https://tinyurl.com/htb8jatf>.
- Cerdán Martínez, V. y Padilla Castillo, G. (2019). Historia del fake audiovisual: deepfake y la mujer en un imaginario falsificado y perverso. *Historia y comunicación social*, 24(2), 505-520. DOI: 10.5209/hics.66293.
- Chesney, R. y Citron, D. (2018, 11 de diciembre). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*. <https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war>.
- Christov, A. (2018). Economics Effects of the Fake news on Companies and The Need of New PR Strategies. *Journal of Sustainable Development*, vol. 8, issue 20, 41-49. <https://www.ceeol.com/search/article-detail?id=714176>.
- Cole, S. (2017, 11 de diciembre). AI-Assisted Fake Porn Is Here and We're All Fucked. *Vice*. https://motherboard.vice.com/en_us/article/gdydm/gal-gadot-fake-ai-porn.
- Comisión Europea. (2018). Plan de Acción contra la desinformación. Comunicación conjunta al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. <https://bit.ly/3LpQbXx>.
- Cruz, E. P. F. da, Pereira, R. M., Jubini, G. M., Quarto, L. C. y de Souza, C. H. M. (2021). Fake news: uma revisão compreensiva e interdisciplinar. *Cadernos de Educação Tecnologia e Sociedade*, 14(3), 502-520. <https://brajets.com/v3/index.php/brajets/index>.
- Dasilva, J. P.; Ayerdi, K. M. y Galdospin T. M. (2021). Deepfakes on Twitter: Which Actors Control Their Spread? *Media and Communication*, vol. 9, issue 1, ISSN: 2183-2439, 301-312. DOI: 10.17645/mac.v9i1.3433.
- Deshmukh A. y Wankhade S.B. (2021). Deepfake Detection Approaches Using Deep Learning: A Systematic Review. In: Balas V. E., Semwal V. B., Khandare A., Patil M. (eds.): *Intelligent Computing and Networking. Lecture Notes in Networks and Systems*, vol. 146. Springer, Singapore. <https://www.springer.com/series/15179>.
- De Veirman, M., Cauberghe, V. y Hudders, L. (2017). Marketing through Instagram influencers: The impact of number of followers and product divergence

- on brand attitude. *International Journal of Advertising*, 36(5), 798-828. DOI: [10.1080/02650487.2017.1348035](https://doi.org/10.1080/02650487.2017.1348035).
- Elías Zambrano, R., Jiménez-Marín, G., y García Medina, I. (2018). Educomunicación, televisión y valores. Análisis de la programación desde una óptica publicitaria. *Educação & Linguagem*, 21(1), 95-107.
- Esteves, F. y Sampaio, G. (2019). *Viral: a epidemia de fake news e a guerra da desinformação*. Porto Salvo, Portugal. Desassossego.
- Estudio de Comunicación e Servimedia. (2018). Influencia de las noticias falsas en la opinión pública. *Servimedia*. <https://bit.ly/3rV7Xul>.
- Europol. (2021). EU Serious and Organised Crime Threat Assessment (SOCTA). https://www.europol.europa.eu/cms/sites/default/files/documents/socta2021_1.pdf.
- European Commission. (2020). Proposal for an AI Regulation laying down harmonised rules on artificial intelligence. COM/2021/206 final. <https://digital-strategy.ec.europa.eu/news-redirect/709090>.
- EXAME. (2024, 12 de mayo). Professores gerados por inteligência artificial dão aulas em universidade de Hong Kong. *EXAME*. <https://tinyurl.com/52etkbwu>.
- Fallis, D. (2015). What Is Disinformation? *Library Trends*, 63(3), 401-426. DOI: [10.1353/lib.2015.0014](https://doi.org/10.1353/lib.2015.0014).
- Fallis, D. (2020). The Epistemic Threat of Deepfakes. *Philosophy & Technology*, 34: 623-643. DOI: [10.1007/s13347-020-00419-2](https://doi.org/10.1007/s13347-020-00419-2).
- Farquhar, M. (2005). *A treasury of deception*. Penguin. New York.
- FBI. (2021). Malicious Actors Almost Certainly Will Leverage Synthetic Content for Cyber and Foreign Influence Operations. <https://www.ic3.gov/Media/News/2021/210310-2.pdf>.
- Feng, J., Zhang, J., Liu, M. y Fang, Y. (Eds). (2021). Biometric Recognition. CCBP 2021. Lecture Notes. *Computer Science*, vol12878. Springer, Cham. DOI: [10.1007/978-3-030-86608-2_38](https://doi.org/10.1007/978-3-030-86608-2_38).
- Fernández, E. G. (2021). Análisis forense de imágenes y videos. *INFOTEC Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación*. México. <https://infotec.repositorioinstitucional.mx/jspui/handle/1027/475>.
- Gaimari, G. (2021). *Inteligencia artificial e impacto en el cibercrimen*. Doctoral dissertation. Universidad de Belgrano-Facultad de Derecho y Ciencias Sociales-Abogacía. <http://repositorio.ub.edu.ar/handle/123456789/9333>.
- Galston, W. (2020, 8 de enero), Is seeing still believing? The Deepfake Challenge to Truth in Politics. *Brookings Institution*. <https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/>.
- Gamir-Ríos, J. y Tarullo, R. (2022). Predominio de las cheapfakes en redes sociales. Complejidad técnica y funciones textuales de la desinformación desmentida en Argentina durante 2020. *adComunica*, 97-118. DOI: [10.6035/adcomunica.6299](https://doi.org/10.6035/adcomunica.6299).
- García-Ull, F. J. (2021). Deepfakes: el próximo reto en la detección de noticias falsas. *Anàlisi: Quaderns de comunicació i cultura*, 64, 103-120. <https://dialnet.unirioja.es/servlet/articulo?codigo=7987666>.
- Giansiracusa, N. (2021). *How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More*. Apress. USA.

- Graves, L. y Cherubini, F. (2016). The rise of fact-checking sites in Europe. *Reuters Institute for the Study of Journalism*. University of Oxford. Oxford.
- Harwell, D. (2019, 4 de septiembre). An artificial-intelligence first: Voice-mimicking software reportedly used in a major theft. <https://tinyurl.com/2wacndyj/>.
- Huijstee, M., Boheemen, P., Das, D., Nierling, L., Jahnel, J., Karaboga, M. y Fatun, M. (2021). *Tackling Deepfakes in European policy*. European Parliament. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU\(2021\)690039](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2021)690039).
- Ireton, C. y Posetti, J. (2018). *Journalism, 'Fake news' and Disinformation – A Handbook for journalism Education and Training*. France. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000265552>.
- Kietzmann, J., Mills, J. A. y Plangger, K. (2021) Deepfakes: perspectives on the future “reality” of advertising and branding. *International Journal of Advertising*, 40: 3, 473-485. DOI: [10.1080/02650487.2020.1834211](https://doi.org/10.1080/02650487.2020.1834211).
- Kugler, M. B. y Pace, C. (2021). Deepfake Privacy: Attitudes and Regulation. *Northwestern Public Law Research Paper, Rev*, 611(3), 611-680. DOI: [10.2139/ssrn.3781968](https://doi.org/10.2139/ssrn.3781968).
- La Voz del Interior (2019). Los «deepfakes» serán la próxima frontera del fraude empresarial desde 2020, según especialistas. Agencia DPA. 28/11/2019. <https://tinyurl.com/yz5cjj7n>
- Lei n.º 27/2021, de 17 de maio. Diário da República. Carta Portuguesa de Direitos Humanos na Era Digital. 17 de maio de 2021. <https://dre.pt/dre/detalhe/lei/27-2021-163442504>.
- Ley Orgánica 1/1982, de 5 de mayo, BOE, sobre protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen. 14 de mayo de 1982. <https://www.boe.es/buscar/doc.php?id=BOE-A-1982-11196>.
- Magallón-Rosa, R. (2019). *'Unfaking News'*. Cómo combatir la desinformación. Pirámide. Madrid.
- Malheiros, N. D. T. (2021). *O Futuro Chegou: Regulamentação Do Uso De Veículos Autônomos No Brasil*. Dissertação de Mestrado. UNIALFA - Centro Universitário Alves Faria. <http://tede.unialfa.com.br/jspui/handle/tede/377>.
- Maldita.es. (2021, 21 de enero). Por qué los 'deepfakes' no son el problema (aún) y sí los 'cheapfakes'. <https://maldita.es/malditobulo/20190904/por-que-los-deepfakes-no-son-el-problema-aun/>.
- Marcos, N. (2024, 22 de marzo). Así se hizo Alba Renai, la 'influencer' virtual que presenta un programa sobre 'Supervivientes': «No ha venido a quitar el trabajo a nadie». *El País*. <https://tinyurl.com/ffp4t6wf>.
- Masood, M., Nawaz, M., Malik, K., Javed A. y Irtaza, A. (2021). Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward. *ArXiv [cs.CR]* abs/2103.00484. <https://arxiv.org/abs/2103.00484>.
- Mut Camacho, M. y Rueda Lozano, A. (2022). Las empresas ante la desinformación. La necesidad de un nuevo enfoque metodológico. *Vivat Academia. Revista de Comunicación*, 155, 113-129. DOI: [10.15178/va.2022.155.e1327](https://doi.org/10.15178/va.2022.155.e1327).
- Nguyen, T. T., Nguyen, Q. V. H., Nguyen, C. M., Nguyen D. T. y Nahavandi, S. (2021). Deep Learning for Deepfakes Creation and Detection: A Survey. *ArXiv:1909.11573v3 [cs.CV]*. 26 apr 2021. <https://arxiv.org/pdf/1909.11573.pdf>.

- O'Donnell, N. (2021). Have we no decency? section 230 and the liability of social media companies for deepfake videos. *University of Illinois Law Review*. University of Illinois College of Law. <https://illinoislawreview.org/print/vol-2021-no-2/have-we-no-decency/>.
- Palomo-Domínguez, I. (2021). Del mito a la viralidad. El caso de la campaña de Cruzcampo que resucitó a Lola Flores. *adResearch: Revista Internacional de Investigación en Comunicación*, 26, 2021, 38-58 DOI: 10.7263/adresic-026-02.
- Paris, B., Donovan, J. (2021). Deepfakes and cheap fakes. Thousand Oaks: Sage (=Data & Society's Media Manipulation research initiative). <https://datasociety.net/library/deepfakes-and-cheap-fakes/>.
- Pennacchio, K. (16 de mayo). Meios de comunicação, faculdades e governos da América Latina experimentam a criação de apresentadores de TV virtuais usando inteligência artificial. *LatAm Journalism Review*. <https://tinyurl.com/2p9jrnpv>.
- PTSOC News. (2021). Deepfakes: uma nova ciberameaça às organizações. Associação DNS.PT. 3ª edição. Dezembro 2021.
- Proposición de Ley Orgánica 122/000011 de 13 de octubre de 2023, de Regulación de las Simulaciones de Imágenes y Voces de Personas Generadas por Medio de la Inteligencia Artificial. Presentada por el Grupo Parlamentario Plurinacional SUMAR. Congreso de los Diputados. XV Legislatura. Serie B. Núm. 23-1. BOCG-Boletín Oficial de las Cortes Generales. <https://www.congreso.es/gl/proposiciones-de-ley>
- Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). <http://data.europa.eu/eli/reg/2024/1689/oj>
- Reglamento (UE) 2022/2065 de 19 de octubre. Diario Oficial de la Unión Europea. Reglamento de Servicios Digitales, del Parlamento Europeo y del Consejo, relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE. L 277/1. <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32022R2065>.
- Romero, J. C. (2021). Ciberseguridad: Evolución y tendencias. *Boletín IEEE*, (23), 460-494.
- Sánchez-Gey Valenzuela, N., Jiménez-Marín, G. y Román-San-Miguel, A. (2022). La responsabilidad social de las empresas audiovisuales: el caso de la productora Indaloymedia. *Prisma Social*, 37, 238-264. <https://revistaprismasocial.es/article/view/4549/5316>.
- Santaella, L. y de Matto Salgado, M. (2021). Deepfake e as consequências sociais da mecanização da desconfiança. *TECCOGS: Revista Digital de Tecnologias Cognitivas*, (23). <https://revistas.pucsp.br/index.php/teccogs/article/view/55981/37929>
- Sensity. (2021). The State of Deepfakes 2020: Updates on Statistics and Trends. <https://sensity.ai/reports/>.
- Silbey, J. y Hartzog, W. (2020). The Upside of Deep Fakes. *Maryland Law Review*, [s. l.], v. 78(4), 960-966, 2019. <https://digitalcommons.law.umaryland.edu/mlr/vol78/iss4/8/>.

- Soriana. (2021, 4 de mayo). Soriana La de Todos Los Mexicanos. #LaDeTodosLosMexicanos. [Vídeo de youtube]. <https://www.youtube.com/watch?v=S8J1bVbhMNY>.
- Sun P., Yuan Yan, Z., Shen, Z., Shi, S. y Dong, X. (2021). Deepfakes Detection Based on Multi Scale Fusion. In: Feng, J., Zhang, J., Liu, M. y Fang, Y. (eds.): Biometric Recognition. CCBR 2021. Lecture Notes. *Computer Science*, 12878. Springer, Cham. DOI: [10.1007/978-3-030-86608-2_38](https://doi.org/10.1007/978-3-030-86608-2_38).
- Tandoc Jr., E., Lim, Z. y Ling, R. (2017). Defining 'Fake news': A Typology of Scholarly Definitions. *Digital Journalism*, 6(3): 1-17. https://www.researchgate.net/publication/319383049_Defining_Fake_News_A_typology_of_scholarly_definitions.
- Valero, P. P. y Oliveira, L. (2018). Fake news: una revisión sistemática de la literatura. *Observatorio (OBS*)*, 12(5). DOI: [10.15847/obsOBS12520181374](https://doi.org/10.15847/obsOBS12520181374).
- Vizoso, Á., Vaz-Álvarez, M. y López-García, X. (2021). Fighting deepfakes: Media and internet giants' converging and diverging strategies against hi-tech misinformation. *Media and Communication*, 9(1), 291-300. DOI: [10.17645/MAC.V9I1.3494](https://doi.org/10.17645/MAC.V9I1.3494).
- Vosoughi, S., Roy, D. y Aral, S. (2018). The spread of true and false news online. *Science*, 359, 1146-1151. DOI: [10.1126/science.aap9559](https://doi.org/10.1126/science.aap9559).
- Volkoff, V. (2000). *Pequena história da desinformação: do cavalo de Tróia à Internet*. Editorial Notícias. Lisboa.
- Wardle, C. y Derakhshan, H. (2017). Information Disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe Report*, vol27, 1-107. Council of Europe. Strasbourg. <https://surl.li/pmyaeb>.
- Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11). DOI: [10.22215/timreview/1282](https://doi.org/10.22215/timreview/1282).
- Wiesenberg, M., Zerfass, A. y Moreno, A. (2017). Big data and automation in strategic communication. *International Journal of Strategic Communication*, 11(2), 95-114. <http://dx.doi.org/10.1080/1553118X.2017.1285770>.
- Wodajo, D. y Atnafu, S. (2021). Deepfake Video Detection Using Convolutional Vision Transformer. ArXiv:2102.11126 [Cs, Cv]. 11 mar 2021. <https://arxiv.org/abs/2102.11126>.
- Xavier, C. (2024, 2 de enero). "Channel 1" será o primeiro canal de notícias sem pivots reais. IA chega ao jornalismo. *CNN Portugal*. <https://cnnportugal.iol.pt/videos/channel-1-sera-o-primeiro-canal-de-noticias-sem-pivots-reais-ia-chega-ao-jornalismo/65946d5f0cf25f99539252ff>.
- Yang, J., Shuai Xiao, S., Li, A., Lan, G. y Wang, H. (2021). Detecting fake images by identifying potential texture difference. *Future Generation Computer Systems*, vol. 125, 127-135. DOI: [10.1016/j.future.2021.06.043](https://doi.org/10.1016/j.future.2021.06.043).
- Zazpe, P. R. (2019). El fenómeno de la desinformación. Análisis crítico y propuestas de actuación desde el ámbito académico (actualizado). *Verdad y falsedad de la información*. UNAM, Instituto de Investigaciones Bibliotecológicas y de la Información, 125-142. <https://eprints.ucm.es/id/eprint/60713/>.

Cómo investigar la inteligencia artificial: tres estudios de caso

Santiago Tejedor

Universitat Autònoma de Barcelona
Santiago.tejedor@uab.cat

Stephanie Vick

Universitat Autònoma de Barcelona
Stephanie.vick@uab.cat

1 | Introducción

La inteligencia artificial (IA) se ha convertido en un tema central en el ámbito del periodismo debido al profundo impacto que ha generado en la producción informativa y en los modelos de negocio de los medios de comunicación. Este fenómeno abarca diversas facetas que incluyen desde los aspectos éticos y las implicaciones sociales hasta su aplicación práctica en las redacciones y la viabilidad de su integración a largo plazo. Ante la complejidad de estos desafíos, resulta fundamental la colaboración entre los medios de comunicación y la academia, con el fin de combinar el conocimiento científico con las realidades y dinámicas del entorno periodístico. En este contexto, la presente investigación analiza tres proyectos de investigación que abordan la aplicación de la IA al periodismo desde diferentes enfoques, con el objetivo de explorar sus posibilidades, retos y el potencial de sus soluciones en el contexto actual.

2 | Revisión de la literatura

En los últimos años, el panorama de los medios de comunicación ha atravesado transformaciones sin precedentes, gracias a los avances significativos de las tecnologías de la información y la comunicación (TIC) que marcan el ritmo de las innovaciones, además de su rol de renovar y modernizar las prácticas comunicativas (Ali & Hassoun, 2019). Sin duda, la tecnología es una fuerza en el sector mediático para producir contenidos que van en línea con las demandas de los usuarios de internet. Por eso, se puede decir que los cambios dramáticos en el periodismo están conectados directamente con los avances de las herramientas tecnológicas (Veglis & Bratsas, 2017).

Los medios alrededor del mundo se enfrentan a muchos retos como consecuencia de los cambios radicales de la industria gracias a la digitalización. Por eso, a diario se puede evidenciar cómo estas empresas de comunicación buscan soluciones para mantenerse al día con las exigencias de un periodismo más tecnológico y moderno (Leppanen *et al.*, 2017).

No debería ser sorpresa que la irrupción de la inteligencia artificial (IA) en la producción de contenidos periodísticos marque otro punto de inflexión en la evolución de los medios de comunicación (Apablaza-Campos & Wilches Tinjacá, 2024). Esta tecnología no solo ha irrumpido en la prensa o con noticias automatizadas textuales, sino que se ha extendido a todas las fases del *newsmaking* (Túñez-López *et al.*, 2021) para jugar un papel integral en el ecosistema mediático actual (Ali & Hassoun, 2019).

Como plantea Calily (2018), el uso de la IA se ha vuelto indispensable. Entre sus posibilidades, puede aumentar la producción de contenidos (Papadimitriou, 2016), personalizar informaciones (Newman *et al.*, 2019), identificar eventos de interés periodístico para su difusión (Steiner, 2014), luchar contra la desinformación (Flew *et al.*, 2012), entre otras cuestiones. Su creciente irrupción está transformando las formas de captar, procesar, generar y distribuir las informaciones y el trabajo de los periodistas (Avilés, 2019).

Tal y como exponen Túñez y Tejedor (2019), se trata de un escenario de cambios tan acelerados, que la investigación desde la academia, en colaboración con diferentes medios y proyectos, resulta decisiva. En la actualidad, el desarrollo de la IA en los procesos industriales es una de las prioridades de los programas de investigación a todos los niveles, y los beneficios anticipados no se limitan a la dimensión económica e industrial (Túñez, Toural & Cacheiro, 2018), sino que se extienden a la esfera social (Peña-Fernández *et al.*, 2023). El libro blanco sobre IA de la Comisión Europea (2020) plantea que su uso permite hacer frente a retos como mejorar la calidad democrática o la prestación de servicios públicos de calidad. La centralidad del servicio a la ciudadanía más

allá de solo el desarrollo tecnológico es lo que debe definir este proceso, con el objetivo de que alcance un crecimiento económico y social sostenible.

En un reporte realizado por el grupo Polis de la Escuela de Economía de Londres (Becket & Yassen, 2023), los resultados apuntan a que la IA ya tiene una presencia significativa en los medios, con el 75 por ciento de editores de diversos medios asegurando que la utilizan en algún aspecto del proceso periodístico con la principal motivación de buscar mayor eficiencia e incrementar la productividad. No obstante, entre las conclusiones principales del informe destaca la desigualdad marcada en el acceso a las herramientas de IA, dependiendo del tamaño del medio y región de origen, y la resistencia de la industria al cambio y el miedo a la pérdida de empleo.

Las investigaciones sobre la colaboración entre humanos y máquinas en las redacciones ayudan a evaluar las posibilidades y limitaciones de la IA en el periodismo, más allá de la simple publicación de noticias, abarcando aspectos como la mejora de calidad informativa y la interacción con los usuarios (Lewis, Guzman & Schmidt, 2019).

En el caso particular de España, hay mucho interés, poca innovación. En un informe de la Academia DataFactory sobre la IA en Iberoamérica, el catedrático de Periodismo e investigador de medios digitales de la Universidad de Navarra Ramón Salaverría apunta a que, aunque los medios españoles sienten un gran interés por la IA, su aproximación sigue siendo cautelosa (Apablaza-Campos *et al.*, 2024). Los medios de comunicación y los periodistas enfrentan el reto tecnológico a través de iniciativas específicas, pero sin emprender transformaciones de gran escala.

La investigación de Calvo-Rubio y Ufarte-Ruiz (2020) proporciona una visión clave sobre el impacto de la IA en el periodismo en España, al analizar la percepción de casi 300 periodistas y expertos. Los resultados plantean que un 88 por ciento considera que la IA ofrece ventajas para la profesión y un 80 por ciento cree que facilitará diversas tareas. Estos desafíos demandan de un trabajo desde la academia (Tejedor *et al.*, 2024a) que ha de adaptar el diseño curricular al nuevo escenario.

Por su parte, el estudio de De Lara, García-Avilés y Arias-Robles (2022) explora cómo se está implementando la IA en áreas específicas del periodismo. Entre los resultados más destacables, su artículo apunta a que la mitad de los expertos entrevistados coincidió en que, en la actualidad, la IA se encuentra en una fase de experimentación. Consideran que se está investigando y probando, pero no existen, por el momento, grandes proyectos que se hayan consolidado. No obstante, los entrevistados mencionan que la IA está siendo utilizada con éxito en España para automatización de contenidos y herramientas destinadas a combatir la desinformación (Tejedor *et al.*, 2024b).

Una de las áreas en las que la IA ha encontrado mayor aceptación es en la lucha contra la desinformación y es ahí donde el periodismo resulta

primordial. Los periodistas necesitan hoy más que nunca soportes tecnológicos para fortalecer sus mecanismos de verificación. Para ello, la IA representa un aliado para el desempeño de sus tareas diarias como en el rol de verificación de contenidos (Tejedor, Cervi & Sancho, 2023). Por ejemplo, los proyectos de la Unión Europea FactCheckEU y FANDANGO sobre desinformación a través de IA y *big data* han obtenido resultados positivos por su capacidad de filtrar datos y crear *softwares* que ayuden a empresas de redes sociales a remover cuentas falsas y detectar desinformación (Moreno Espinosa, Abdulasam Alsarayreh & Figuereo-Benítez, 2024).

La relación entre IA y periodismo se ha profundizado, no solo por la producción periodística, sino también por la perspectiva académica de conscientemente educar a los profesionales en el uso de estas herramientas (Pavlik, 2023). Las posibilidades de la IA en los medios de comunicación se ha convertido en un tema que genera gran interés por sus implicaciones industriales y sociales (Tuñez, Toural & Cacheiro, 2018). Por eso, es importante destacar aquellos proyectos que están a la vanguardia de la inclusión de IA y ponerlos a disposición de los medios y la ciudadanía.

3 | Metodología

Este trabajo tiene como objetivo reflexionar sobre tres proyectos de inteligencia artificial aplicada al periodismo. Para alcanzar este objetivo, se ha utilizado la técnica del estudio de casos (Yin, 1994), un método empírico que se centra en el análisis de fenómenos contemporáneos dentro de su contexto real, permitiendo así una investigación intensiva de uno o más casos específicos. Este enfoque promete proporcionar información valiosa que puede extenderse a una población más amplia de casos, contribuyendo al desarrollo de conocimientos más generales sobre el fenómeno en cuestión (Gerring, 2017).

Entre los diferentes tipos de estudio propuestos por Stake (2005), esta investigación se enfoca en el estudio de caso colectivo, que se realiza cuando el interés de la investigación se centra en un fenómeno, población o condición general, seleccionando para ello varios casos que se han de estudiar.

A partir de este acercamiento, se analizarán tres proyectos de IA aplicada al periodismo. Estos proyectos fueron seleccionados con base a su relevancia, representatividad y el impacto que han tenido o tendrán en la industria periodística a partir de la aplicación de IA.

Utilizando un acercamiento cualitativo, se evaluarán las propuestas, metodologías, objetivos y entregables de los proyectos seleccionados, con el fin de identificar su efectividad en la implementación de soluciones en entornos reales de comunicación y periodismo.

Los estudios cualitativos son más amigables de adaptar al estudio de caso, ya que tratan de temas que son únicos y que tienen determinadas

características que ameritan un estudio profundo y un acercamiento más real del contexto donde se desarrolla el fenómeno a investigar (Jiménez Chávez & Comet Weiler, 2016).

4 | **Discusión y resultados**

El Gabinete de Comunicación y Educación de la Universitat Autònoma de Barcelona (UAB) es un grupo consolidado y especializado en la investigación y divulgación científica, que pertenece al Departamento de Periodismo y Ciencias de la Comunicación de la UAB. Reconocido por AGAUR (Agència de Gestió d'Ajuts Universitaris i de Recerca) de la Generalitat de Catalunya como grupo de investigación consolidado en función de su trayectoria, proyección y desarrollo, desarrolla proyectos e investigaciones en el terreno de convergencia entre la comunicación y la educación.

Desde su creación, el Gabinete de Comunicación y Educación ha desarrollado iniciativas destinadas a integrar las tecnologías de la comunicación en la denominada sociedad global o del conocimiento.

Entre sus diferentes iniciativas, el Gabinete de Comunicación y Educación ha desarrollado y colaborado en proyectos relacionados a la inteligencia artificial (IA), la *media literacy* y la desinformación. Entre ellos, destacan tres proyectos que se encuentran a la vanguardia de estos tópicos y prometen ser referentes en el contexto iberoamericano.

4.1. **Proyecto IVERES**

El proyecto IVERES «Identificación, Verificación y Respuesta. El Estado democrático ante el reto de la desinformación interesada» es un proyecto financiado por el Ministerio de Ciencia e Innovación perteneciente a la convocatoria de 2021, Proyectos I+D+I en líneas estratégicas, en colaboración público-privada, del programa estatal I+D+I orientada a los retos de la sociedad en el marco del plan estatal de investigación científica y técnica y de innovación 2017-2020.

Codirigido por Radio Televisión Española (RTVE) y la Universitat Autònoma de Barcelona (UAB), el proyecto tiene como objetivo desarrollar, mediante herramientas de IA, un sistema automático o semiautomático de verificación de contenidos de audio, vídeo, imagen y redes sociales que se adapte a las necesidades de los periodistas.

RTVE tiene como misión el respeto a la veracidad y el rigor de sus contenidos, lo que exige adaptar esta misión al presente para disponer de mecanismos en los que los bulos y la desinformación no penetren en la difusión del medio y para que la ciudadanía disponga de un lugar fiable para comprobar los contenidos.

La lucha contra la desinformación se plantea como una materia de seguridad nacional de alta prioridad, y en este contexto, el principal objetivo del proyecto IVERES es desarrollar un sistema de verificación en lengua española. En el marco del proyecto, se ha creado una herramienta integral que proporciona un amplio abanico de funcionalidades para facilitar el trabajo de los profesionales y se ha establecido una base sólida para futuras innovaciones en el campo de la verificación de contenidos mediante herramientas de IA.

El sistema de verificación permite abordar las siguientes prioridades temáticas:

1. La actividad política institucional, por ser fuente habitual de campañas de desinformación interesadas y por el potencial de desestabilizar el sistema democrático de libertades.
2. La actualidad científica y médica, por ser fuente habitual de campañas de desinformación interesada y por su capacidad de ser dañinas para un gran número de personas.
3. La actualidad deportiva, por la relativa facilidad tecnológica que supone desarrollar una herramienta con actores y temáticas delimitables.
4. La polarización y los discursos de odio.
5. La igualdad entre sexos.
6. Las migraciones.

Para poder desarrollar el aplicativo, el proyecto se llevó a cabo en una serie de fases que permitieron detectar las necesidades reales del trabajo, conocer las herramientas disponibles y su eficacia y llevar a cabo un trabajo previo para el posterior diseño del flujo de trabajo y de la interfaz de verificación. Asimismo, cada fase desarrolló una serie de metodologías que hicieron posible abordar los acercamientos necesarios para el desarrollo de la herramienta (tab. 12.1).

Las primeras fases del proyecto fueron dirigidas a identificar y establecer una red de contactos con grupos de empresas y entidades involucradas en el ámbito de la verificación de noticias. Esto permitió recopilar información sobre el sector en general, sus necesidades técnicas, herramientas disponibles y posibles colaboraciones. En paralelo, se realizaron investigaciones y estudios específicos sobre herramientas tecnológicas para la verificación en los diferentes ámbitos profesionales.

El trabajo previo de investigación le permitió al equipo diseñar un modelo experimental de la herramienta que se fue actualizando según el desarrollo del proyecto hasta lograr el aplicativo que actualmente está disponible en la web. El desarrollo tecnológico estuvo a cargo de tres universidades públicas españolas: la Universidad Carlos III de Madrid, Universidad de Granada y la Universidad Politécnica de Cataluña.

Tabla 12.1. Plan de trabajo del proyecto IVERES según objetivos, metodología y entregables

Objetivo	Metodología	Entregable
Tejer una red de confianza con otras instituciones públicas y privadas que se sientan concernidas por los objetivos del proyecto	Recopilación bibliográfica de <i>stakeholders</i> Entrevistas Reflexión y preparación de protocolos	Documentación de entrevistas Listado de <i>stakeholders</i> Protocolo entre organizaciones Listado herramientas de verificación Sistematización del marco teórico
Evaluar distintos paquetes de verificación existentes en el entorno anglosajón y emitir un diagnóstico crítico sobre las posibilidades de adaptación en castellano	Recopilación de actores relevantes, buenas prácticas y herramientas Establecimiento de catálogo de herramientas de referencia Diseño de modelo experimental	Documentación entrevistas de actores relevantes Catálogo de herramientas Diseño de modelo experimental
Establecer un modelado completo de la herramienta	Diseño ontológico de capa de IA Prototipado de diseño web	Diseño ontológico Prototipo de la web interfaz de usuario
Desarrollar el primer prototipo	Programación de capas de IA Desarrollo de entorno web	Herramienta de verificación de vídeo Herramienta de verificación de audio Herramienta de monitorización de redes sociales, viralización de contenido y archivado
Aplicar la herramienta en un entorno controlado de verificación	Aplicación de la herramienta en entornos controlados Corregir errores detectados	Informes sobre aplicación de herramientas desarrolladas
Aplicar la herramienta en un entorno de redacción real	Aplicación de la herramienta Corregir errores detectados	Informe sobre aplicación de herramientas desarrolladas
Aplicar conocimientos obtenidos para proponer metodología de aplicación y uso	Redacción de un protocolo de uso de las herramientas	Protocolos de uso
Abrir a un grupo tasado de instituciones la herramienta, validar su uso y posibles mejoras	Grupos focales cualitativos Informe de los grupos focales Publicaciones científicas	Informe de grupos focales

Fuente. Elaboración propia

Primeramente, la herramienta permite la monitorización de vídeos y canales en redes sociales como Youtube, TikTok, X y Telegram. En el caso de los vídeos, permite detectar automáticamente el idioma, obteniendo transcripciones y traducciones en más de 96 idiomas y ofrece metadatos y subtítulos de utilidad para el verificador. Además, permite la generación automática de correos de alerta y clasificación de contenidos de los canales monitoreados.

En adición, mediante diferentes algoritmos de IA, la herramienta permite realizar verificación de vídeos e imágenes, arrojando intervalos de probabilidad de que sean falsos o no.

La plataforma IVERES también es capaz de verificar y detectar posibles contenidos falsos e identificar cuentas en redes sociales que podrían ser bots, lo cual es crucial para combatir la desinformación.

Además, se desarrollaron métodos para discernir entre grabaciones de audio reales y suplantaciones automatizadas, una capacidad esencial en la era de noticias falsas y la manipulación de voz.

Por otro lado, el sistema monitoriza redes sociales y sigue cuentas sospechosas de propagar información falsa, y hace seguimiento de viralización de contenidos en redes. Por último, el aplicativo crea, mantiene y actualiza ontologías referidas a los tres objetivos del proyecto: salud, política y deporte.

Desde sus primeras fases, IVERES hizo esfuerzos comunicativos y campañas para influir en el debate público, especialmente la comunidad periodística, para mostrar tanto las potencialidades como las limitaciones de las herramientas de IA para la verificación de contenidos.

Actualmente, el servicio del aplicativo está disponible para los periodistas y la ciudadanía en general.

4.3. Observatorio para la Innovación de los Informativos en la Sociedad Digital

El Observatorio para la Innovación de los Informativos en la Sociedad Digital (OI2) es un proyecto impulsado por la cátedra de Radio Televisión Española (RTVE) y la Universitat Autònoma de Barcelona, puesto en marcha en 2015. Este proyecto se centra en el análisis exhaustivo, el monitoreo y la investigación de las tendencias más recientes que están marcando la evolución de los informativos televisivos. Su propósito fundamental es evidenciar cómo estas tendencias emergentes pueden ser adaptadas e implementadas de manera eficaz y sostenible, teniendo en cuenta los desafíos del contexto actual, tanto en términos económicos como sociales.

En el entorno mediático contemporáneo, los medios se enfrentan a retos cada vez más complejos. Uno de los principales desafíos consiste

en atraer nuevas audiencias y ofrecer contenidos innovadores que respondan de manera efectiva a las demandas de un público cada vez más diverso y exigente. Esta necesidad de adaptación y reinención obliga a los medios a explorar nuevas formas de producción y distribución de contenidos, así como a fomentar una mayor interacción con los usuarios.

La experimentación con nuevos formatos y enfoques narrativos representa una oportunidad para los medios de revitalizar los contenidos informativos ante la necesidad de mantenerse relevantes en un entorno mediático saturado. Este proceso de transformación implica, además, una reflexión crítica sobre las metodologías tradicionales del periodismo, así como una exploración de las oportunidades que ofrecen las nuevas tecnologías para renovar las formas de consumir información.

En este contexto, OI2 se erige como una plataforma crucial para estudiar, analizar y aplicar las tendencias más relevantes dentro del ámbito mediático. Su objetivo es proporcionar un marco teórico y práctico que permita explorar las nuevas posibilidades que ofrecen las tecnologías emergentes, particularmente aquellas relacionadas con la digitalización y la IA, para llevar a cabo una práctica periodística más alternativa, eficiente e innovadora.

En el marco de este proyecto, se han desarrollado hasta la fecha seis informes especializados, elaborados por la Cátedra RTVE-UAB, centrados en el uso de IA aplicada a los informativos. Estos informes han abordado de manera detallada el impacto de la IA en diversas fases del proceso comunicativo, desde la recopilación de datos hasta la presentación de contenidos, con el objetivo de optimizar la producción. A través de estas investigaciones, se han identificado múltiples oportunidades que la IA ofrece a los medios de comunicación, con el fin de mejorar su eficiencia operativa, personalizar sus contenidos y beneficiar tanto al medio como a su audiencia (tab. 12.2).

El primer informe, *Detección de noticias a través de aplicaciones de inteligencia artificial*, está dedicado a la detección de noticias desde la primera fase exploratoria que abarca desde que se produce un hecho noticiable hasta que los periodistas se enteran del suceso y deciden cubrirlo.

El informe se guio por la pregunta: ¿puede ser útil la inteligencia artificial a la hora de proporcionar a los periodistas indicios e informaciones sobre hechos singulares que no estén previstos de antemano? Para responderla, la investigación contempla cinco fases relacionadas con la recopilación de información, la elaboración de noticias, la personalización del contenido, el archivo del material y la interacción y participación de la audiencia.

Concretamente, el estudio marcó el panorama de las aplicaciones de IA en el periodismo, identificó cómo estas pueden incorporarse a las redacciones audiovisuales y cuál sería su impacto y estableció un mapa sobre las tendencias de futuro en el sector.

Tabla 12.2. Informes desarrollados por OI2

Informe	Tema	Metodología
<p data-bbox="216 353 462 420">Detección de noticias a través de aplicaciones de inteligencia artificial</p>  <p data-bbox="236 629 443 660">La inteligencia artificial aplicada a informativos 2019-2020</p> <p data-bbox="220 687 455 706">OI² <small>Observatorio para la Innovación de las Tecnologías de la Información y la Comunicación</small> rtve + UMB <small>Universidad del País Vasco</small></p>	<p data-bbox="499 302 784 378">Detección de noticias a través de aplicaciones de inteligencia artificial</p>	<p data-bbox="802 302 1018 378">Revisión de literatura Búsqueda documental Estudio de casos</p>
<p data-bbox="236 789 443 839">Generación automática de textos periodísticos</p>  <p data-bbox="236 1057 443 1088">La inteligencia artificial aplicada a informativos 2019-2020</p> <p data-bbox="220 1115 455 1133">OI² <small>Observatorio para la Innovación de las Tecnologías de la Información y la Comunicación</small> rtve + UMB <small>Universidad del País Vasco</small></p>	<p data-bbox="499 729 752 778">Generación automática de textos periodísticos</p>	<p data-bbox="802 729 1018 857">Revisión de literatura Búsqueda documental Elaboración de marco conceptual Estudio de casos</p>
<p data-bbox="194 1208 488 1270">Personalización de contenidos en medios audiovisuales</p>  <p data-bbox="236 1466 443 1497">La inteligencia artificial aplicada a informativos 2019-2020</p> <p data-bbox="194 1525 488 1556">OI² <small>Observatorio para la Innovación de las Tecnologías de la Información y la Comunicación</small> rtve + UMB <small>Universidad del País Vasco</small></p>	<p data-bbox="499 1157 777 1206">Personalización de contenidos en medios audiovisuales</p>	<p data-bbox="802 1157 1018 1233">Revisión de literatura Búsqueda documental Estudio de casos</p>

Informe	Tema	Metodología
<p>Contra la desinformación: Estrategias y herramientas</p>  <p>La inteligencia artificial aplicada a informativos 2019-2020</p> <p>OI² Observatorio para la Formación de los Informadores de la Sociedad Digital. rtve + UMB Universidad de Barcelona</p>	<p>Contra la desinformación, estrategias y herramientas</p>	<p>Revisión de literatura Búsqueda documental Estudio de casos</p>
<p>Asistentes y Presentadores Virtuales en el Periodismo</p>  <p>La inteligencia artificial aplicada a informativos 2019-2020</p> <p>OI² Observatorio para la Formación de los Informadores de la Sociedad Digital. rtve + UMB Universidad de Barcelona</p>	<p>Asistentes y presentadores virtuales en el periodismo</p>	<p>Revisión de literatura Búsqueda documental Estudio de casos</p>
<p>La aplicación de la IA a los sistemas de edición</p> <p>2023-2024</p> <p>Santiago Tejedor & Pere Vila (Dir.)</p> <p>OI² Observatorio para la Formación de los Informadores de la Sociedad Digital. rtve UAB Universidad de Barcelona</p>	<p>La aplicación de la IA a los sistemas de edición</p>	<p>Revisión de literatura Búsqueda documental Estudio de casos</p>

Fuente. Elaboración propia

En el caso del segundo informe *Generación automática de textos periodísticos*, se dedicó a estudiar la escritura automática de textos a través del uso de inteligencia artificial. El escrito se centró en recopilar conocimiento científico sobre la cuestión, el desarrollo industrial y las aplicaciones comerciales existentes, las perspectivas de futuro y las consecuencias de la aplicación de la IA en el trabajo periodístico.

De esta forma, la investigación se sustenta de una base bibliográfica científica y académica vinculada al tema y construye un marco conceptual que gira en torno al proceso de generación automática de textos periodísticos. Además, el informe presenta varios casos de herramientas, describe las tendencias y analiza las repercusiones y consecuencias en el trabajo de los periodistas.

El tercer informe *Personalización de contenidos en medios audiovisuales* se dedicó a investigar sobre la personalización de contenidos en medios audiovisuales a través de herramientas de IA. La investigación se centró en explorar las posibilidades, oportunidades, beneficios y eventuales perjuicios de la aplicación de estos sistemas.

Entre los contenidos, el proyecto explora el concepto de personalización, partiendo de las conceptualizaciones del marketing, para a través de ellas proporcionar un modelo que dé cuenta de las posibles aplicaciones de la personalización en el campo del periodismo. Luego, identifica los principales servicios y herramientas del mercado y se realiza un estudio de casos.

En el caso del informe cuatro *Contra la desinformación, estrategias y herramientas*, se centra en las nuevas herramientas de verificación y el uso de IA para contrarrestar la desinformación. En el marco de este informe, se abordó el tema de la desinformación, incluyendo iniciativas internacionales, sistemas de *fact checking* y herramientas.

El documento incluye una síntesis sobre el concepto de desinformación desde una perspectiva europea y explica los tres tipos de sistemas de verificación: periodística, colaborativa y automatizada. Luego, aborda los hallazgos científicos e incluye un listado de iniciativas de alcance internacional y *fact-checkers* habituales. Por último, ofrece un estudio de caso de medios que utilizan herramientas de verificación.

Por su parte, el quinto informe sobre *Asistentes y presentadores virtuales en el periodismo* se enfoca en los asistentes y presentadores virtuales desarrollados para el periodismo. Mediante una investigación exhaustiva, se analiza la evolución, cambios y transformaciones de los avances de este campo.

El informe hizo hincapié en la investigación sobre los desarrollos de la IA vinculados al aprendizaje profundo, la generación de lenguaje natural, las tecnologías de realidad extendida o virtual y la producción de medios sintéticos que interactúan con humanos. El estudio identifica nuevas líneas de investigación sobre tecnologías de asistentes y presentadores

virtuales en el periodismo y examina el factor ético asociado a estas tecnologías.

Por último, el sexto informe, *La aplicación de la IA a los sistemas de edición*, explora la integración de la IA en ámbitos de la producción de contenidos audiovisuales como imagen, vídeo y texto, destacando cómo esta tecnología puede transformar los métodos tradicionales de creación y distribución de contenidos.

Se trata de un directorio de herramientas, proyectos y empresas que aplican la IA a los sistemas de edición de vídeo, imagen y texto. Además, se evalúa el impacto de la IA en los sistemas, considerando tanto los desafíos tecnológicos como las oportunidades para innovar en la forma en la que se crean, se distribuyen y consumen.

En conclusión, los informes presentados y abiertos a la audiencia ofrecen un análisis integral sobre el impacto de la IA en diversas facetas del periodismo. Este trabajo recurrente destaca cómo la IA está transformando el periodismo, ofreciendo tanto oportunidades innovadoras como desafíos éticos y tecnológicos que deben ser considerados para asegurar una integración efectiva y responsable. La cátedra continúa en la elaboración de otros informes que permitirán ampliar la investigación en IA y promover la apuesta por el uso de IA para reformar el periodismo.

4.4. La inteligencia artificial para el fomento del periodismo de calidad y la *media literacy*: avances tecnológicos aplicados y desafíos en la era de la desinformación (IA-COM)

El proyecto «La inteligencia artificial para el fomento del periodismo de calidad y la alfabetización mediática: Avances tecnológicos aplicados y desafíos en la era de la desinformación» (Referencia PID2023-149759OB-I00) es una iniciativa derivada de la convocatoria 2023 de «Proyectos de generación de conocimiento» y actuaciones para la formación de personal investigador predoctoral asociadas a dichos proyectos, en el marco del Programa Estatal para Impulsar la Investigación Científico-Técnica y su Transferencia, del Plan Estatal de Investigación Científica, Técnica y de Innovación 2021-2023. Aunque aún se encuentra en su fase inicial, se prevé un impacto a nivel internacional por los objetivos y entregables que se discutirán a continuación.

El proyecto parte de la necesidad de que el ecosistema mediático en España y la región iberoamericana implemente directrices, estrategias y hojas de ruta que fomenten el uso de IA para mejorar los procesos de producción y contenidos informativos y la promoción de nuevos contenidos y servicios.

IA-COM tiene como objetivo identificar las posibilidades de la IA para el fomento de un periodismo de calidad que impulse la *media literacy* y potencie la mejora del ecosistema mediático a partir de transformaciones en el proceso productivo. En adición, prevé elaborar un libro blanco para promover el periodismo de calidad y el pensamiento crítico desde la alfabetización mediática en la era de IA a partir del estudio de los avances tecnológicos aplicados y los retos que plantea en la era de la desinformación. Los resultados pretenden ser difundidos y transferidos a los sectores científicos, profesionales y sociales mediante una serie de formatos y entregables disruptivos.

El proyecto tiene un carácter descriptivo y exploratorio, apoyado en un trabajo metodológico híbrido entre lo cualitativo y cuantitativo. La siguiente tabla muestra las diferentes tecnologías que se utilizarán en los diferentes objetivos y entregables pautados (tab. 12.3).

Tabla 12.3. Plan de trabajo de IA-COM según objetivos, metodología y entregables

Objetivo	Metodología	Entregable
Elaborar un libro blanco para promover el periodismo de calidad y el pensamiento crítico desde la alfabetización mediática en la era de la IA basado en el estudio de los avances tecnológicos aplicados y los retos que plantea la era de la desinformación	Grupos de discusión	1 libro blanco
Analizar los últimos avances publicados en investigación científica sobre a soluciones integradas de IA aplicadas al periodismo	Revisión sistemática de literatura científica	Revisión de literatura científica
Construir una colección de 30 casos de estudio y buenas prácticas a nivel iberoamericano sobre la implementación de la IA en el periodismo para promover la alfabetización mediática	Estudio de casos	30 estudios de casos
Elaborar una cartografía multimedia interactiva (mapeo) de proyectos, iniciativas y buenas prácticas periodísticas con IA	Estudio de casos Cartografía multimedia	1 cartografía multimedia
Analizar los planteamientos, recomendaciones y reflexiones de 50 expertos nacionales e internacionales de diferentes ámbitos y disciplinas del conocimiento sobre la convergencia entre periodismo, IA, alfabetización mediática y desinformación	Entrevistas en profundidad	50 entrevistas en profundidad
Promover la sinergia entre grupos y equipos de investigación de centros, institutos y facultades de las universidades españolas y, por extensión, de la región iberoamericana	Grupos de discusión	6 grupos de discusión

Fuente: elaboración propia

La metodología propuesta se basa en el estudio de la producción científica en el área de convergencia entre periodismo, alfabetización mediática e IA. En este sentido, como elemento novedoso se destaca la combinación de técnicas diversas y la lectura final desde un enfoque híbrido (cuantitativo y cualitativo). Además, se pone énfasis en procesos que permitan generar entregables disruptivos y efectivos para impactar a los destinatarios del proyecto. Además de estos entregables, el proyecto plantea un proceso de difusión en diferentes redes y organizaciones mediáticas para lograr un impacto en la ciudadanía más allá del mundo académico y periodístico.

IA-COM propone aportar soluciones a diferentes problemas existentes relacionados con la transformación digital y su utilización en el periodismo. Además, pretende dar respuesta a las necesidades de tres líneas de trabajo: inclusión de cadenas de valor de IA, inclusión social de la IA y la transformación digital de sectores productivos.

En el tema de la inclusión de cadenas de valor de IA, el proyecto busca identificar nuevos modelos de negocio en periodismo basados en IA, adoptar nuevos desarrollos tecnológicos en el trabajo periodístico, acelerar los procesos de inclusión de IA en el periodismo, impulsar e identificar experiencias de automatización y nuevos roles y perfiles profesionales a partir de estas dinámicas.

En el caso de la inclusión social de IA, se promoverá un periodismo de calidad basado en las posibilidades de la IA, se fomentará la alfabetización mediática de periodistas y medios de comunicación para la lucha contra la desinformación, se establecerán códigos deontológicos para el periodismo que use IA y se dará visibilidad a problemas relacionados con la brecha de género, brecha digital o el medio ambiente.

Por último, sobre la transformación digital de sectores productivos, el proyecto prevé idear nuevos productos informativos, estudiar nuevos formatos de comunicación a partir de la IA, establecer sinergias entre sectores involucrados, aplicación de dinámicas de procesamiento de lenguaje natural y aprovechamiento del internet en el trabajo periodístico.

Los resultados del proyecto tendrán un impacto tanto científico como técnico. La propuesta permitirá mejorar la organización, rutinas de producción y generación de contenidos en los medios gracias a los procesos de trabajo predictivos, de automatización y robotización que la IA posibilitará. De esta forma, se espera un avance técnico inexistente hasta la fecha en la transferencia del potencial de la IA en el escenario mediático.

En el caso del impacto científico, los resultados obtenidos en las diferentes fases de investigación serán compartidos en revistas especializadas en periodismo y/o ingeniería y en congresos nacionales e internacionales.

IA-COM representa un hito en la inclusión responsable de la IA en los medios de comunicación para beneficio de la sociedad y la lucha contra la desinformación.

Conclusiones

Los proyectos analizados en este estudio reflejan un panorama positivo en relación con la investigación, transferencia de conocimiento y aplicabilidad de la IA en el ámbito del periodismo. Cada una de estas iniciativas ha demostrado estar alineada con las necesidades actuales del sector periodístico, abordándolas desde perspectivas diversas: OI2 centrado en un enfoque técnico y teórico y los proyectos IVERES e IA-COM adoptando un acercamiento más práctico y orientado a la implementación directa de soluciones.

El proyecto IVERES representa un avance crucial en la integración de IA como herramienta para combatir la desinformación. Este sistema, diseñado en español, tiene la ventaja de ser adaptable al contexto iberoamericano, una región que con frecuencia se ve limitada por el predominio de herramientas y sistemas orientados al inglés. Esta característica permite a IVERES ofrecer una opción más contextualizada y eficaz a las particularidades de los medios hispanohablantes. Además, el proyecto establece una base sólida para futuras innovaciones gracias al potencial de expansión que ofrece y la apuesta por el sector público como propulsor de nuevas estrategias para mejorar los contenidos y renovar los procesos de producción.

Por su parte, OI2 destaca por su impacto considerable en la investigación sobre IA, al ofrecer una recopilación exhaustiva de las actualizaciones más relevantes sobre esta tecnología emergente. Este proyecto subraya las oportunidades del uso de la IA y presenta una visión integral del estado de la cuestión, abordando tanto las propuestas teóricas como las aplicaciones prácticas. Al proporcionar información científica de vanguardia y herramientas prácticas, OI2 contribuye a comprender cómo la IA puede transformar el ejercicio periodístico, al mismo tiempo que actúa como puente entre la academia y la práctica profesional.

Por último, IA-COM se configura como un paso esencial hacia un periodismo más eficiente, responsable y ético, apoyado en tecnologías de IA. A través de la implementación de tecnologías emergentes, este proyecto busca optimizar la calidad de la producción informativa, enfrentar los desafíos asociados con la desinformación y fomentar la alfabetización mediática. La combinación de innovación tecnológica con un enfoque ético y responsable tiene el potencial de redefinir los estándares del periodismo en un mundo cada vez más influenciado por la inteligencia artificial.

En cuanto a las metodologías empleadas por los tres proyectos, es importante resaltar que la integración de enfoques tanto cuantitativos como cualitativos ha sido clave para la generación de entregables que abarcan diversas facetas de la investigación. Los métodos cuantitativos han sido útiles para delinear el panorama general, como, por ejemplo, al identificar la cantidad de herramientas existentes y la adopción de IA

en medios de comunicación. Sin embargo, este enfoque no podría ser considerado exhaustivo sin el complemento de metodologías cualitativas, que permiten profundizar en las características específicas de los fenómenos que se están estudiando. Las metodologías cualitativas, en particular, son fundamentales para ofrecer una visión detallada y enriquecedora, permitiendo explorar las implicaciones y dinámicas del uso de la IA en el periodismo

En conclusión, los tres proyectos representan avances significativos en la investigación y aplicación de la IA en el ámbito periodístico, especialmente en el contexto iberoamericano. Si bien cada uno de ellos aborda aspectos distintos del fenómeno, juntos constituyen un paso crucial hacia la transformación del sector, contribuyendo a una mayor eficiencia, responsabilidad y ética en la producción informativa. En términos de futuras líneas de investigación, sería pertinente explorar el impacto de la IA en la formación y capacitación de periodistas, así como su papel en la evolución de los modelos de negocio en los medios de comunicación, considerando también los aspectos éticos y sociales asociados con su integración en el periodismo.

Referencias

- Ali, W. y Hassoun, M. (2019). Artificial Intelligence and Automated Journalism: Contemporary Challenges and New Opportunities. *International Journal of Media, Journalism and Mass Communications*, 5(1). <https://doi.org/10.20431/2454-9479.0501004>.
- Apablaza-Campos, A., Wilches Tinjacá, J. A. y Salaverría, R. (2024). Generative Artificial Intelligence for Journalistic Content in Ibero-America: Perceptions, Challenges and Regional Projections. *BiD*, 52. [10.1344/bid2024.52.06](https://doi.org/10.1344/bid2024.52.06).
- Apablaza-Campos, A. y Wilches Tinjacá, J. A. (2024). Inteligencia artificial para la generación de contenidos en Iberoamérica: experiencias editoriales en medios de comunicación. *DataFactory, Institución Universitaria Politécnico Granacolombiano & Iniciación Científica*. <https://doi.org/10.15765/librosic.v5i160>.
- Calvo-Rubio, L. M. y Ufarte-Ruiz, M. J. (2020). Percepción de docentes universitarios, estudiantes, responsables de innovación y periodistas sobre el uso de inteligencia artificial en periodismo. *Profesional de la Información*, 29(1). <https://doi.org/10.3145/epi.2020.ene.09>.
- Chaves, V. E. J. y Weiler, C. C. (2016). Los estudios de casos como enfoque metodológico. *DOAJ (DOAJ: Directory Of Open Access Journals)*. <https://doaj.org/article/9bbc23caf2a54afc8166f8ad5f2d9a0b>.
- Comisión Europea (2020). *Libro blanco sobre la inteligencia artificial. Un enfoque europeo orientado a la excelencia y la confianza*. <https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1>.
- Flew, T., Spurgeon, C., Daniel, A. y Swift, A. (2012). The promise of computational journalism. *Journalism Practice*, 6(2). <https://doi.org/10.1080/17512786.2011.616655>.

- Galily, Y. (2018). Artificial intelligence and sports journalism: Is it a sweeping change? *Technology In Society*, 54, 47-51. <https://doi.org/10.1016/j.techsoc.2018.03.001>.
- Gerring, J. (2007). Case study research: principles and practices. *Choice Reviews Online*, 45(03), 45-1530. <https://doi.org/10.5860/choice.45-1530>.
- Leppänen, L., Munezero, M., Granroth-Wilding, M. y Toivonen, H. (2017, 1 de enero). *Data-Driven news generation for automated journalism*. <https://doi.org/10.18653/v1/w17-3528>.
- Lewis, S. C., Guzman, A. L. y Schmidt, T. R. (2019). Automation, Journalism, and Human-Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News. *Digital Journalism*, 7(4), 409-427. <https://doi.org/10.1080/21670811.2019.1577147>.
- Newman, N., Fletcher, R., Kalogeropoulos, A. y Nielsen, R. K. (2019). Reuters Institute digital news report 2019. *SSRN Electronic Journal*, 2019. <https://apo.org.au/node/241541>.
- Papadimitriou, A. (2016). The Future of Communication: Artificial Intelligence and Social Networks. *Media & Communication Studies*. Malmö University. <http://bit.ly/379xa70>.
- Ruiz, M. J. U., Rubio, L. M. C. y Verdú, F. J. M. (2021). Los desafíos éticos del periodismo en la era de la inteligencia artificial. *Estudios Sobre el Mensaje Periodístico*, 27(2), 673-684. <https://doi.org/10.5209/esmp.69708>.
- Stake, R. E. (2005) Investigación con estudio de casos. Madrid, Morata. https://edmorata.es/wp-content/uploads/2022/06/STAKE.InvestigacionEstudioCasos_prw-1.pdf.
- Steiner, T. (2014, 17 marzo). *Telling Breaking News Stories from Wikipedia with Social Multimedia: A Case Study of the 2014 Winter Olympics*. [arXiv.org](https://arxiv.org/abs/1403.4289). <https://arxiv.org/abs/1403.4289>.
- Tejedor, S., Cervi, L., Romero-Rodríguez, L. M., Vick, S. (2024). Integrating Artificial Intelligence and Big Data in Spanish Journalism Education: A Curricular Analysis. *Journal Media*, 5, 1607-1623. <https://doi.org/10.3390/journalmedia5040100>.
- Tejedor, S., Cervi, L., Pulido, C. y Guerrero, N. (2024). Inteligencia Artificial para creación de contenido periodístico: Un benchmarking de las principales herramientas. *SOCIAL REVIEW. International Social Sciences Review Revista Internacional De Ciencias Sociales*, 12(1), 15-30. <https://doi.org/10.37467/revsocial.v12.5179>.
- Túñez-López, J., Fieiras-Ceide, C. y Vaz-Álvarez, M. (2021). Impact of Artificial Intelligence on Journalism: transformations in the company, products, contents and professional profile. *Communication & Society*, 34(1), 177-193. <https://doi.org/10.15581/003.34.1.177-193>.
- Túñez-López, J. y Tejedor-Calvo, S. (2019). Inteligencia artificial y periodismo [presentación del monográfico]. *Doxa Comunicación Revista Interdisciplinaria de Estudios de Comunicación y Ciencias Sociales*, 29, 163-168. <https://doi.org/10.31921/doxa-com.n29a8>.
- Veglis, A. y Bratsas, C. (2017). Towards A Taxonomy of Data Journalism. *Journal Of Media Critiques*, 3(11), 109-121. <https://doi.org/10.17349/jmc117309>.
- Yin, R. K. (1984). *Case Study Research: Design and Methods*. <https://cds.cern.ch/record/1171670>.

Autorías

■ Barceló, Yasir

Universidad de Málaga

yasir@uma.es

Doctorando en la Universidad de Málaga, donde investiga sobre la inteligencia artificial en medios de comunicación. Actualmente es miembro del equipo de trabajo de los proyectos DesinfoperIA: Aplicaciones periodísticas de la IA para reducir la desinformación: tendencias, usos y percepciones de profesionales y audiencias; DesinfoAND: Impacto de la desinformación en Andalucía: Análisis transversal de las audiencias y las rutinas y agendas periodísticas, y gen-IA: Biblioteca de herramientas de inteligencia artificial para la creación de contenidos en medios de comunicación. Su trabajo se centra en analizar el impacto de estas tecnologías en la creación y difusión de contenidos.

■ Benítez-Eyzaguirre, Lucía

Universidad de Cádiz

Lucia.benitez@gm.uca.es

Profesora titular de Comunicación Audiovisual y Publicidad de la Universidad de Cádiz. Coordinadora del Máster en Gestión Estratégica e Innovación en Comunicación. Doctora por la Universidad de Sevilla. Premio Extraordinario de Doctorado de la Universidad de Sevilla y Premio RTVA a la mejor tesis doctoral de Comunicación. Máster en Tecnologías digitales y sociedad del conocimiento. Lucía Benítez-Eyzaguirre investiga la comunicación digital y los usos de la tecnología, desde una perspectiva de género.

■ Blanco, Sonia

Universidad de Málaga

sblanco@uma.es

Profesora titular del departamento de Comunicación Audiovisual y Publicidad de la Universidad de Málaga. Principales líneas de investigación: inteligencia artificial generativa, desinformación, redes sociales, periodismo digital, *podcasting*, audiencias activas. Es investigadora de los proyectos DesinfoPerIA: Aplicaciones periodísticas de la IA para reducir la desinformación: tendencias, usos y percepciones de profesionales y audiencias; DesinfoAND: Impacto de la desinformación en Andalucía: Análisis transversal de las audiencias y las rutinas y agendas periodísticas. Además de pertenecer a estos proyectos nacionales y autonómicos, actualmente lidera el proyecto gen-IA: Biblioteca de herramientas de inteligencia artificial para la creación de contenidos multimedia. <https://gen-ia.uma.es/>.

■ Bringué Sala, Xavier

Universidad de Navarra

jbringue@unav.es

Licenciado en Ciencias de la Educación, doctor en Comunicación y máster en Matrimonio y Familia por la Universidad de Navarra. Es profesor titular en Comunicación Audiovisual y Publicidad en dicha universidad. Ha dirigido múltiples proyectos de investigación en *marketing*, jóvenes y medios de comunicación para diversas instituciones. Fundador y expresidente del Consejo Académico del Foro Generaciones Interactivas, también es miembro del Observatorio del Juego Infantil.

■ Calvo Olloqui, Enma

Universidad Complutense de Madrid

encalvo@ucm.es

Investigadora y docente en la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid, donde cursa un doctorado en Comunicación Audiovisual enfocado en el cine dirigido por mujeres en España, con una beca FPU. Forma parte del Grupo de Estudios Cinematográficos (ESCINE) y el proyecto de innovación docente Laboratorio transmedia. Es becaria de investigación en la Residencia de Estudiantes de Madrid. Graduada en Comunicación Audiovisual (USJ) y máster en Escritura Creativa (UCM), ha recibido las becas Heraldo de Aragón y Excelencia. Su cortometraje (2019) ganó el premio a Mejor Ópera Prima en el Festival de Cine de Fuentes.

■ Cantón Correa, Francisco Javier

Universidad Internacional de la Rioja

javicanton@ugr.es / javier.canton@unir.net

Investigador y profesor en las Universidades de Granada, UNIR, UOC e Isabel I. Sociólogo y comunicador, ha sido periodista en RTVE, dentro de la unidad Verifica RTVE. Actualmente investiga sobre la relación entre IA y desinformación dentro del proyecto IBERIFIER.

■ Cea, Nereida

Universidad de Málaga

nereidacea@uma.es

Profesora titular en el departamento de Periodismo de la Universidad de Málaga. Doctora en Ciencias de la Información en la Universidad de Navarra. Sus líneas de investigación giran en torno al estudio de los mercados de la comunicación y la economía de los medios, la desinformación y la IA. Ha impartido docencia en la Universidad Internacional de Catalunya y en la Universitat Oberta de Catalunya, entre otras. En el ámbito profesional, trabajó como redactora en *La Verdad* (Grupo Vocento), *La Economía* (Grupo Intereconomía), *Cinco Días* (Prisa) y *El Mundo* (Unidad Editorial). Fue responsable de comunicación de la Cámara de Comercio de Murcia y en la Consejería de Presidencia de la comunidad autónoma de la Región de Murcia.

■ Deltell Escolar, Luis

Universidad Complutense de Madrid

deltell@ucm.es

Catedrático en Comunicación Audiovisual. Cineasta y autor de diversos artículos de investigación. Fue coordinador del doctorado en Comunicación Audiovisual y Publicidad y coordinador del Grado en Comunicación Audiovisual. Codirige el grupo complutense de investigación ESCINE sobre estudios cinematográficos. Su campo de investigación es la historia de la imagen, el cine y el análisis de las nuevas tecnologías de las ciencias de la información. Además, ha dirigido documentales y cortometrajes con los que ha logrado una veintena de premios internacionales.

■ Gomes Gonçalves, Sonia

Universidad da Madeira

soniaggoncalves17@gmail.com

Investigadora en Comunicación. Posee un doctorado internacional en Comunicación por la Universidad de Sevilla, con mención de cum laude, y un máster en *Lingüística por la Universidad Nova de Lisboa, con especialización en terminología del periodismo*. Cuenta con una larga experiencia en el campo de la comunicación. Periodista de profesión, trabaja actualmente y desde hace más de veinte años, en el Gobierno regional

de Madeira, en áreas de asesoría de comunicación. Igualmente, su perfil docente e investigador la ha llevado a impartir clases de publicidad, relaciones públicas y *marketing* desde hace más de 15 años.

■ Huertas Bailén, Amparo

Universitat Autònoma de Barcelona

amparo.huertas@uab.cat

Profesora titular en el departamento de Comunicación Audiovisual y Publicidad de la Universitat Autònoma de Barcelona (UAB). Directora del Instituto de la Comunicación (InCom-UAB). Responsable del Grupo Internacional de Estudios sobre Comunicación y Cultura (2021SGR-00179). Coordinadora de la Cátedra UNESCO de Comunicación. Directora de la revista *Quaderns del CAC*.

■ Lleida Solano, Eduardo

Universidad de Zaragoza

lleida@unizar.es

Doctor ingeniero de Telecomunicación por la Universidad Politécnica de Cataluña y catedrático de universidad en el área de Teoría de la Señal y Comunicaciones. Es miembro del Instituto de Investigación en Ingeniería de Aragón, donde lidera el grupo de investigación Vivolab, dedicado fundamentalmente a la investigación y desarrollo en tecnologías del habla y el lenguaje y el aprendizaje automático. Es director de la Cátedra RTVE en la Universidad de Zaragoza «hacia la automatización de la documentación de contenidos audiovisuales y sonoros».

■ Niño Romero, Mónica

Universidad de Huelva

monica.nino@alu.uhu.es

Doctoranda en Comunicación por las universidades de Málaga, Sevilla, Huelva y Cádiz (línea en Educomunicación y Alfabetización Mediática, *Media Literacy*). Licenciada en Periodismo, experta universitaria en información medioambiental, en *community manager*, redes sociales y *marketing*. Con experiencia como periodista en comunicación institucional y comunicación política, en medios de comunicación, en agencia de publicidad y comunicación digital. Docente en alfabetización mediática y digital. Premio Meridiana 2013 por la Junta de Andalucía por el censo de expertas. Por la empresa GTO, premio Agripina 2016 por la campaña de publicidad.

■ Pereira Correia, Pedro Álvaro

Universidade da Madeira

pacorreia@staff.uma.pt

Profesor e investigador de *Marketing* y Comunicación en la Universidad de Madeira, posee un doctorado en *Marketing* y Comunicación de la Universidad de Vic en España, con mención de cum laude y doctorado europeo. Enseña cursos de eMarketing, Emprendimiento, Gestión de Tecnología de la Información, Recursos Humanos y proyectos en la Universidad de Madeira y en ISAL. Ha supervisado tesis, participado en jurados de doctorado y máster en varios países y organizado seminarios internacionales. Su experiencia profesional incluye puestos como director de marketing y consultor de gestión. Sus intereses de investigación abarcan el marketing y de IA, digital y móvil.

■ Peres-Neto, Luis

Universitat Autònoma de Barcelona

luiz.peres@uab.cat

Profesor lector Serra Húnter en el departamento de Periodismo y Ciencias de la Comunicación. Miembro del Instituto de la Comunicación (In-Com-UAB), donde forma parte del Grupo Internacional de Estudios sobre Comunicación y Cultura (2021SGR-00179)

■ Ramírez-Alvarado, María del Mar

Consejo Audiovisual de Andalucía.

mariam.ramirez.alvarado@juntadeandalucia.es

Consejera del Consejo Audiovisual de Andalucía. Catedrática del área de Comunicación Audiovisual y Publicidad adscrita a la Facultad de Comunicación de la Universidad de Sevilla, de la que ha sido decana.

■ Román-San-Miguel, Aránzazu

Universidad de Sevilla

arantxa@us.es

Profesora titular de universidad en el área de Periodismo en la Facultad de Comunicación de la Universidad de Sevilla, con más de cien publicaciones entre capítulos de libro, artículos de revistas y compilaciones. Directora del Grupo de Investigación HUM-849 Estrategias de Comunicación y del Máster en Periodismo de Datos e Inteligencia Artificial de la Universidad de Sevilla. Su área de especialización se centra en el diseño periodístico, periodismo deportivo y especializado, tecnologías de la información e inteligencia artificial.

■ Sádaba Chalezquer, Charo

Universidad de Navarra

csadaba@unav.es

Catedrática y decana de la Facultad de Comunicación de la Universidad de Navarra. Tras obtener su doctorado, su investigación se ha centrado en el impacto de la digitalización en los modelos de negocio de las empresas de medios. También en los últimos veinte años ha investigado sobre la relación entre niños, niñas, adolescentes y la tecnología. Ha formado parte de numerosos proyectos de investigación en este ámbito, en América Latina, España y Europa. Más recientemente se ha interesado por la competencia digital y la alfabetización mediática en el contexto de la lucha contra la desinformación, como parte del equipo de Iberifier, hub ibérico del EDMO. En 2022 fue parte del grupo de expertos de la Comisión Europea que trabajó en «Tackling disinformation and promoting digital literacy through education and training». Es también co-IP del proyecto Juventud en Transición del ICS de la Universidad de Navarra.

■ Sánchez González, Hada M

Universidad de Sevilla

misago@us.es

Profesora titular en Periodismo de la Universidad de Sevilla, dirige el Laboratorio de Proyectos en Comunicación (LabProCom), el grupo de investigación GIATI - Análisis y Técnica de la Información y la revista científica «Textual & Visual Media» editada por la Sociedad Española de Periodística (SEP). Especializada en periodismo digital, verificación de la información, sistemas de inteligencia artificial, innovación, emprendimiento y social media. Investigadora del proyecto «Impacto de la desinformación en Andalucía: análisis transversal de las audiencias y las rutinas y agendas periodísticas. Desinfoand» Proyectos de Excelencia. ProyExcel_00143 y de gen-IA: Biblioteca de herramientas de inteligencia artificial para la creación de contenidos en medios de comunicación. Ref: JA.B3-23. II Plan Propio de Investigación, Transferencia y Divulgación Científica. Universidad de Málaga. IP del proyecto Inteligencia artificial, realidad mixta y alfabetización en desinformación. Fundación de Investigación (FIU) de la Universidad Sevilla. Muy recientemente, también ha sido investigadora del proyecto I+D+i «Impacto de la desinformación en el periodismo: contenidos, rutinas profesionales y audiencias» del Ministerio de Ciencia e Innovación de España.

■ Sánchez González, María

Universidad de Málaga

m.sanchezgonzalez@uma.es

Profesora asociada doctora en Periodismo de la Universidad de Málaga y técnico en Innovación de la Universidad Internacional de Andalucía.

Acreditada como profesora titular por ANECA (diciembre de 2020). Actualmente es investigadora del proyecto PID2019-108956RB-I00 (Ministerio de Ciencia e Innovación de España); de «MEDIO: Media & Data Innovation» (SEJ 612) de la Junta de Andalucía; de MAEDU —Marco de Análisis del Emprendimiento Digital en las Universidades (Referencia: PID2022-139332NB-I00)—, financiado por el Ministerio de Ciencia e Innovación; y de gen-IA: Biblioteca de herramientas de inteligencia artificial para la creación de contenidos en medios de comunicación. Ref: JA.B3-23. Financiado por el II Plan Propio de Investigación, Transferencia y Divulgación Científica. Universidad de Málaga. Muy recientemente también ha sido investigadora del proyecto de I+D+i «Impacto de la desinformación en el periodismo: contenidos, rutinas profesionales y audiencias (DESINFOPER)». Líneas de investigación: contenidos digitales; social media; audiencias activas, experiencia de usuario y participación ciudadana; innovación; emprendimiento digital; inteligencia artificial, y verificación periodística. También viene trabajando sobre innovación y tecnología educativa, como investigadora y gestora, además de responsable de numerosos proyectos educativos.

■ Sánchez Hunt, Marta

Centro Universitario EUSA Sevilla

marta.hunt@usa.es

Profesora de universidad privada en el grado de publicidad y relaciones publicas en el centro universitario EUSA Sevilla adscrito a la Universidad de Sevilla. Doctora en Comunicación por la Universidad de Málaga. Su área de investigación se centra en las redes sociales, inteligencia artificial, comunicación política, estructuras de poder y marketing digital.

■ Tejedor, Santiago

Gabinete de Comunicación y Educación, Universitat Autònoma de Barcelona

Santiago.tejedor@uab.cat

Catedrático de Periodismo. Dirige el Gabinete de Comunicación y Educación, grupo de investigación reconocido y consolidado de la UAB. Investiga sobre periodismo, narrativas digitales y tecnología educativa.

■ Vick, Stephanie

Gabinete de Comunicación y Educación, Universitat Autònoma de Barcelona

m.gcpe@uab.cat

Máster en Periodismo e Innovación en Contenidos Digitales por la Universidad Autónoma de Barcelona y Grado en Información y Periodismo por la Universidad de Puerto Rico. Posee una amplia experiencia colaborando en diversos medios de comunicación en Puerto Rico y Barcelona. Investigadora del Gabinete de Comunicación y Educación de la UAB.

El rápido avance de la Inteligencia Artificial (IA) ha supuesto un importante cambio de paradigma en el ámbito de la comunicación y de la difusión de información. Su irrupción y rápida evolución, junto con la extendida transformación digital, está generando beneficios de distinto orden. Sin embargo, al mismo tiempo son diversos los riesgos asociados, en especial en lo que respecta a la desinformación.

Los capítulos que conforman esta obra colectiva giran, precisamente, en torno a la IA desde múltiples perspectivas. Por una parte, se aborda su marco jurídico y las consideraciones éticas, sesgos y problemáticas, así como las aplicaciones de IA para abordar los desafíos sociales, formas de exclusión y ciberseguridad. Por otro lado, se incluyen también trabajos centrados en la desinformación, los contextos tecnológicos y digitales en los que esta se produce y el impacto en la opinión pública o las medidas para contrarrestarla. De igual manera, se ofrecen en esta publicación enfoques metodológicos de interés para el estudio de la IA y la desinformación y textos sobre perspectivas educativas, consumo, defensa de las audiencias y tecnologías emergentes en la difusión de la información y otros entornos. Por último, los capítulos incluidos conectan la IA y la desinformación en cuanto a desafíos éticos e impacto en el periodismo y la comunicación audiovisual.

Este volumen, que aparece en el catálogo de la Editorial Universidad de Sevilla, surge del interés del Consejo Audiovisual de Andalucía por conocer y difundir investigaciones científicas. También de la voluntad por transferir conocimiento a la sociedad y por conectar con la academia y con los centros universitarios en los que se llevan a cabo investigaciones sobre el audiovisual y materias relacionadas. Así, el CAA se esfuerza por dar seguimiento y promocionar las aportaciones de la producción científica sobre temas de su competencia.