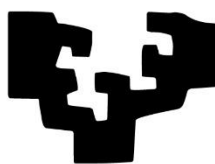


eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

El impacto de la Inteligencia Artificial en los medios de comunicación vascos: Gobernanza algorítmica, soberanía lingüística y el desafío de la integridad informativa

Barbara Sarrionandia Uriguen
-Tesis doctoral-

Simón Peña Fernández
Jesús Ángel Pérez Dasilva
-Dirección-



“La inteligencia consiste no sólo en el conocimiento, sino también en la destreza de aplicar los conocimientos en la práctica”

- Aristóteles

Agradecimientos.....	4
Resumen.....	5
Laburpena	6
Abstract	7
1. INTRODUCCIÓN	8
1.1. Contextualización general del tema	9
1.2. Justificación de la unidad temática.....	12
1.3. Ejes transversales de la investigación.....	14
1.4. Estructura general de la tesis	18
2. Objetivos	22
2.1. Objetivo general	23
2.2. Objetivos específicos.....	24
2.3. Correspondencia entre objetivos y artículos	25
3. Hipótesis de investigación	29
3.1. Hipótesis principal	30
3.2. Hipótesis secundarias.....	30
4. Marco teórico	33
4.1. Principios y aproximaciones teóricas.....	34
4.2. Conceptos y definiciones	37
4.3. Estado de la cuestión.....	39
5. Metodología.....	44
5.1. Diseño de investigación.....	45
6. RESULTADOS	50
7. Conclusiones y discusión	64
9. Limitaciones del trabajo	71
10. Propuestas para futuras líneas de investigación	74
11. Anexos	77
11.1.Trabajos publicados.....	78
11.2. Anexo. Programa formativo: Gobernanza Ética y Soberanía Tecnológica en el Periodismo Vasco	179
11.3. Anexo. Marco normativo para una inteligencia artificial ética y responsable en los medios de comunicación del País Vasco	182
12. Cláusula final.....	185
BIBLIOGRAFÍA	186

Índice de tablas y figuras

Figura 1: ROI global frente la implementación en el ecosistema mediático vasco.....	9
Figura 2: Estructura de la IA en los medios del País Vasco	13
Tabla 3: Artículos publicados de la tesis por compendio:	17
Tabla 4: Correspondencia entre los principales ejes temáticos de la tesis doctoral y las publicaciones del compendio	18
Tabla 5: Organización detallada de la tesis por bloques y contenidos.....	19
Figura 6: Objetivo general de la investigación	23
Figura 7: Objetivos específicos de la investigación	25
Figura 9: Hipótesis principal de la investigación	30
Figura 10: Hipótesis secundarias de la investigación	32
Tabla 12: Registro de entrevistas en profundidad (Informantes clave)	46
Tabla 13: Matriz de técnicas aplicadas por artículo.....	48
Tabla 14: Matriz de triangulación de hallazgos, evidencias y producción científica	51
Figura 15: Análisis de resultados	61

Agradecimientos

La realización de esta tesis doctoral ha sido posible gracias al apoyo, la orientación y la confianza de algunas personas a las que deseo expresar en estas líneas mi más sincero agradecimiento.

En primer lugar, quiero manifestar mi profunda gratitud a mis directores de tesis, Simón Peña Fernández y Jesús Ángel Pérez Dasilva, cuyo acompañamiento ha sido decisivo a lo largo de todo este proceso. Su rigor académico, su generosidad intelectual y su apoyo constante han marcado de manera determinante el desarrollo de esta investigación. Sus observaciones, consejos y estímulos han sido una guía inestimable que ha permitido que este trabajo alcance la coherencia, la solidez y la calidad que aspiraba a tener.

Asimismo, deseo agradecer de manera especial a mi familia, verdadero pilar de este recorrido. A ellos debo no solo el ánimo y la paciencia en los momentos más exigentes, sino también el sentido último de este esfuerzo. Esta tesis se ha construido para ellos y gracias a ellos: por su apoyo incondicional, por estar presentes en cada paso y por recordarme siempre la importancia de mantener el equilibrio entre la dedicación académica y la vida personal.

Extiendo también mi agradecimiento a todas las personas que, de una u otra manera, han contribuido al avance de este proyecto: colegas investigadores, profesionales de los medios de comunicación (Vocento, EITB, Berria, Grupo Noticias, Ser Euskadi, Naiz) que participaron en entrevistas y encuestas, y amistades que brindaron su aliento en los momentos clave. Sin su colaboración y respaldo, este camino habría sido mucho más difícil de transitar.

A todas y todos, gracias.

Resumen

Esta tesis doctoral examina de manera crítica el impacto de la inteligencia artificial (IA) en el periodismo, desde una perspectiva situada y atenta a las particularidades del ecosistema mediático del País Vasco. Lejos de los marcos globales que suelen describir la IA desde experiencias anglófonas y centralizadas, este trabajo analiza cómo estas tecnologías se introducen, se regulan y se discuten en un entorno regional, multilingüe y con una identidad mediática propia. El compendio está formado por seis artículos —tres empíricos, dos teórico-críticos y uno comparativo— que, en conjunto, trazan un panorama amplio y coherente sobre los efectos de la IA en las rutinas profesionales, los marcos éticos, la organización interna de los medios y la representación social y de género. La metodología integra una encuesta a 501 periodistas, entrevistas a responsables tecnológicos y estudios de caso en Eitb y Cadena SER Euskadi, lo que permite observar tanto tendencias generales como procesos concretos de adopción tecnológica. Los resultados muestran que la incorporación de IA en el periodismo vasco es aún temprana y desigual. Persisten brechas formativas, carencias normativas y limitaciones asociadas a la disponibilidad de recursos, especialmente en euskera. Al mismo tiempo, surgen experiencias que revelan un potencial narrativo y verificativo significativo cuando la tecnología se integra bajo criterios de responsabilidad y control editorial. En el plano conceptual, la tesis propone la idea de “apropiación contextualizada” para describir cómo la IA se adapta, negocia y resignifica en función de factores culturales, lingüísticos y organizativos. En el plano práctico, se ofrecen recomendaciones para fortalecer la gobernanza ética, mejorar la alfabetización mediática y promover políticas participativas que garanticen transparencia y equidad. Así, la investigación aporta una mirada crítica, situada y propositiva que permite comprender cómo la IA reconfigura el periodismo en un contexto regional y multilingüe, y sugiere caminos para una innovación tecnológica más democrática y sostenible.

Laburpena

Tesi honek modu kritikoan aztertzen du adimen artifizialak (AA) kazetaritzan duen eragina, ikuspegi kokatu batetik eta Euskal Herriko komunikazio-ekosistemaren berezitasunei erreparatuz. Testuinguru anglofono eta zentralizatueta oinarritutako ikuspegi orokorretatik aldenduta, lan honek aztertzen du teknologia horiek nola txertatzen diren, nola arautzen diren eta nola eztabaidatzen diren ingurune erregional, eleanitz eta nortasun mediatiko bereziko batean. Konpendioa sei artikuluk osatzen dute —hiru ikuspegi enpirikokoak, bi ikuspegi teoriko-kritikokoak eta bat konparatiboa— eta, guztien artean, AAren eraginak modu integratuan mapatzen dituzte: lan-errutinetan, etika-markoetan, erakundeen antolaketan eta gizarte- eta genero-adierazpenetan. Metodologiak 501 kazetariri egindako inkesta, eragile teknologikoei egindako elkarrizketak eta EITB zein Cadena SER Euskadiko kasu-azterketak uztartzen ditu, bai joera orokorrak bai adopzio teknologikoko prozesu zehatzak identifikatzeko aukera emanez. Emaiztekin erakusten dute AAren erabilera oraindik goiztiarra eta desorekatua dela euskal kazetaritzan. Heziketa-hausturak, arau-gabeziak eta baliabideen eskuragarritasunarekin lotutako mugak ageri dira, bereziki euskararen kasuan. Aldi berean, badira esperientzia berritzaileak, AA narrazio- eta egiaztatze-tresna gisa erabil daitekeen potentziala erakusten dutenak, betiere erantzukizun- eta kontrol-irizpide argien pean txertatzen denean. Ikuspegi kontzeptualetik, tesiak “egokitzapen testuinguratuaren” ideia proposatzen du: AA faktore kultural, linguistiko eta antolaketakoen arabera nola negoziatzen, egokitzen eta berrinterpretatzen den azaltzeko. Ikuspegi praktikotik, gobernantza etikoa indartzeko, alfabetatze mediatikoa hobetzeko eta gardentasuna eta ekitatea bermatuko dituzten politika parte-hartzaileak sustatzeko gomendioak egiten dira. Horrenbestez, ikerketa honek ikuspegi kritiko, kokatu eta proposatiboa eskaintzen du, AAk kazetaritza nola eraldatzen duen ulertzeko testuinguru erregional eta eleanitz batean, eta berrikuntza teknologikoa modu demokratikoago eta iraunkorragoan bideratzeko bideak iradokitzen ditu.

Abstract

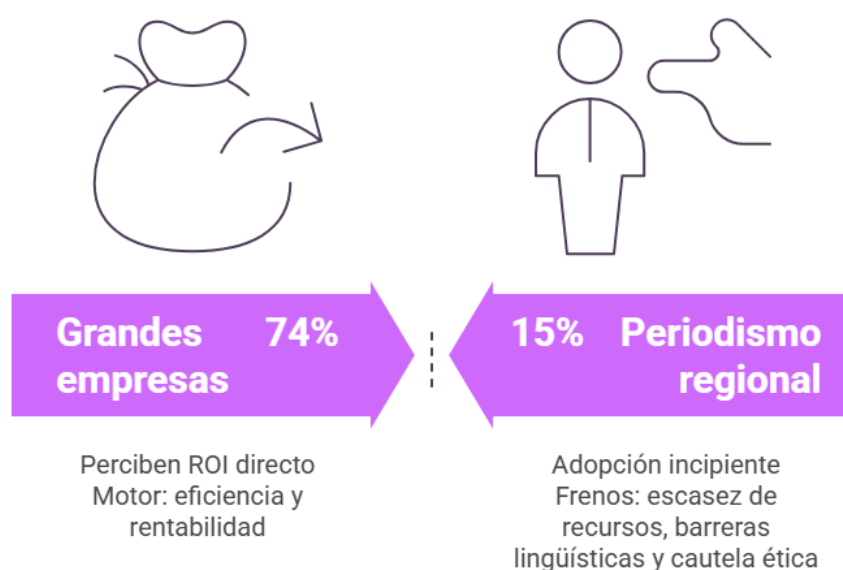
This doctoral dissertation critically examines the impact of artificial intelligence (AI) on journalism from a situated perspective, attentive to the specific characteristics of the media ecosystem in the Basque Country. Moving away from global frameworks that often describe AI through anglophone and centralized experiences, this work analyzes how these technologies are introduced, regulated, and debated within a regional, multilingual environment with its own media identity. The compendium consists of six articles —three empirical, two theoretical-critical, and one comparative— which together outline a broad and coherent picture of AI's effects on professional routines, ethical frameworks, newsroom organization, and social and gender representation. The methodology integrates a survey of 501 journalists, interviews with technological leads, and case studies in EitB and Cadena SER Euskadi, enabling the observation of both general trends and concrete processes of technological adoption. The findings show that the incorporation of AI in Basque journalism remains early and uneven. Training gaps, regulatory shortcomings, and limitations linked to resource availability —particularly in Basque— persist across the sector. At the same time, innovative experiences emerge that reveal significant narrative and verification potential when technology is integrated under principles of responsibility and strong editorial oversight. At the conceptual level, the dissertation proposes the notion of “contextualized appropriation” to describe how AI is adapted, negotiated, and re-signified according to cultural, linguistic, and organizational factors. At the practical level, it offers recommendations to strengthen ethical governance, improve media literacy, and promote participatory policies that guarantee transparency and equity. Overall, the research provides a critical, situated, and forward-looking lens that helps illuminate how AI is reshaping journalism in a regional and multilingual context, and it points toward pathways for more democratic and sustainable technological innovation.

1. INTRODUCCIÓN

1.1. Contextualización general del tema

En el escenario mediático contemporáneo, la integración progresiva de tecnologías de inteligencia artificial (IA) ha pasado de ser una actualización técnica a un cambio estructural que redefine la práctica informativa. A nivel global, este fenómeno se inserta en un contexto de digitalización acelerada, fragmentación del consumo y presiones económicas crecientes que obligan a las redacciones a redefinir sus procesos de producción y distribución (Ioscote et al., 2024; Kanasheva et al., 2024). No obstante, esta transformación no se despliega de forma homogénea, sino que convive con desigualdades estructurales profundas entre regiones. La literatura científica evidencia cómo las diferencias en infraestructuras, el acceso a datos masivos y la capacidad de capacitación técnica condicionan la apropiación tecnológica, creando brechas significativas entre el norte global y los contextos periféricos (Adjin-Tettey et al., 2024; Adefioye, 2024; Kioko et al., 2022). Esta asimetría se ve reforzada por la denominada 'paradoja de la adopción': mientras informes de Google Cloud (2024) señalan que el 74% de las grandes empresas ya perciben un retorno de inversión (ROI) directo tras integrar IA generativa en sus procesos de producción, en el periodismo regional esta cifra cae drásticamente tal como se aprecia en la comparativa de la Figura 1. Esto evidencia que la IA en los medios de proximidad no sigue una lógica de mercado puro, sino una lógica de responsabilidad pública condicionada por la escasez de recursos específicos.

Figura 1: ROI global frente la implementación en el ecosistema mediático vasco



Fuente: Elaboración propia basada en The ROI of Generative AI (Google Cloud, 2024) y los resultados de la encuesta a profesionales del sector.

Por otro lado, esta expansión de la IA automatiza rutinas e interviene en la redacción de contenidos, dando lugar a perfiles profesionales híbridos como el *journalist-programmer*, una figura que sintetiza las tensiones entre la innovación técnica y la necesidad de mantener un control humano estricto basado en competencias analíticas y éticas (Pinto & Barbosa, 2024).

Esta reconfiguración de los roles y procesos no solo depende de la capacitación humana, sino fundamentalmente de la naturaleza de la propia tecnología, que actúa, ante todo, como una infraestructura mediada por el lenguaje. Es en este punto donde el impacto de la IA presenta un matiz crítico, pues la arquitectura de estos sistemas no es neutra respecto al idioma en el que operan. La mayoría de los modelos de lenguaje de gran escala se desarrollan en inglés y otras lenguas dominantes, limitando la presencia en el desarrollo algorítmico de idiomas como el euskera, el gallego o el quechua (Cenoz, 2008; García-Pablos et al., 2024; Salas-Pilco et al., 2022). Bajo esta lógica, la capacidad de una redacción para innovar queda subordinada a la riqueza de su ecosistema lingüístico digital. Se comprende, por tanto, que la desigualdad lingüística se traduce en una desigualdad informativa, ya que la escasez de recursos robustos y la falta de *datasets* específicos empobrecen la visibilidad de las narrativas locales o, en el peor de los casos, las invisibilizan por completo. Es precisamente en este cruce entre la automatización, los valores profesionales y la pluralidad lingüística donde se sitúa la relevancia de analizar casos situados, demostrando que las ventajas y los riesgos de la IA no son universales, sino profundamente dependientes de las condiciones sociotécnicas de cada territorio.

Bajo esta premisa, el caso del País Vasco (Euskadi) se erige como un escenario de análisis excepcional para observar la compleja interacción entre lengua, identidad y tecnología. Con una población de aproximadamente 2,24 millones de habitantes y un régimen de cooficialidad lingüística amparado por el artículo 6 del Estatuto de Autonomía, la región articula un sólido compromiso institucional con la diversidad que se proyecta sobre un ecosistema mediático de notable complejidad. Sin embargo, esta voluntad política convive con una realidad técnica desafiante: aunque el 62,4% de la ciudadanía posee conocimientos de euskera (Eustat, 2021), su peso en la producción informativa diaria aún debe sortear importantes obstáculos estructurales para alcanzar una presencia plena en la era digital.

Esta singularidad lingüística se sustenta sobre una arquitectura industrial de gran calado. Según datos del Observatorio Vasco de la Cultura (2023), el sector de los medios de comunicación y la publicidad en Euskadi aglutina a más de 700 empresas que generan un valor añadido bruto superior a los 500 millones de euros anuales, lo que evidencia una densidad empresarial superior a la media de otras comunidades autónomas. Dicha estructura se vertebra sobre la coexistencia de tres ejes fundamentales: el servicio público, los grandes grupos multimedia privados y un capilar tejido de medios de proximidad. En el primer eje, el grupo Euskal Irrati Telebista (EiTB) no solo ejerce como referente informativo con un presupuesto que

supera los 180 millones de euros, sino que actúa como el principal motor de innovación y tracción tecnológica del territorio. Paralelamente, la hegemonía del mercado privado se concentra en grupos como Vocento —cuya influencia a través de cabeceras históricas como *El Correo* y *El Diario Vasco* ostenta una penetración superior al 60% de los lectores de prensa en sus respectivos territorios (AMI, 2023)— y el Grupo Noticias, que garantiza la vertebración territorial con diarios como *Deia* o *Noticias de Gipuzkoa*. A este mapa se añade un elemento diferencial clave: el tejido de medios locales y comarcales, muchos de ellos integrados en la asociación Hekimen, que resultan esenciales para garantizar la producción informativa en euskera y la cohesión social en el ámbito local.

Sin embargo, la solidez de este sector no garantiza por sí sola una transición tecnológica uniforme. Los datos obtenidos en la presente investigación corroboran que la IA está penetrando en las redacciones vascas de manera silenciosa, situándose todavía en una fase de presencia cauta y fragmentada. De hecho, cerca de la mitad de los periodistas encuestados declara no utilizar estas herramientas en su práctica diaria, señalando que el 63,91% identifica las barreras éticas como el freno principal para su integración. A esta reticencia axiológica se suma una brecha formativa crítica, dado que solo un 14% de los profesionales ha recibido instrucción específica, dependiendo en su gran mayoría del autoaprendizaje. Este déficit formativo limita severamente la capacidad del sector para aplicar, evaluar o supervisar de manera crítica estas tecnologías.

A pesar de estas limitaciones, comienzan a emerger señales de cambio y experimentación. Medios de referencia como EITB o Cadena SER Euskadi ya han iniciado proyectos piloto, como la clonación de voz, aunque siempre bajo protocolos de supervisión editorial estricta. Este patrón de comportamiento confirma una de las tesis centrales de este trabajo: más que una sustitución de la labor humana, la IA está abriendo un espacio de negociación donde los medios exploran la utilidad instrumental de la tecnología sin renunciar a los valores esenciales del oficio, tales como la veracidad, la autoría y la responsabilidad social (Gutiérrez-Caneda et al., 2023; Lopez et al., 2022). En última instancia, las entrevistas realizadas reflejan que el periodismo vasco habita hoy una transición marcada por una mezcla de interés estratégico y cautela operativa, avanzando hacia un horizonte tecnológico que, por el momento, carece de un mapa regulatorio claro y compartido. Esta preocupación por la integridad informativa no es un fenómeno aislado del entorno vasco; el Reuters Institute (2024) subraya que el público global mantiene una 'sospecha generalizada' sobre el uso de la IA en noticias de actualidad (hard news), prefiriendo que los humanos mantengan el control editorial. En este sentido, la necesidad de una 'supervisión humana estricta' que revela esta tesis no es solo una demanda profesional, sino un sello de garantía indispensable para retener la confianza de una audiencia cada vez más escéptica ante lo sintético.

La investigación demuestra que la IA opera como un fenómeno de doble uso en el contexto vasco: posee el potencial de amplificar la desinformación, pero también ofrece recursos avanzados para combatirla, siempre que medie la formación específica y la colaboración entre periodistas, universidades y actores tecnológicos. En Euskadi, la falta de herramientas lingüísticas específicas genera una preocupación palpable entre los profesionales, quienes observan cómo la carencia de modelos adaptados al contexto local puede erosionar la calidad de la información regional. Por ello, esta tesis examina de manera crítica y empírica cómo se apropia la IA en el periodismo vasco, articulando seis dimensiones clave: ética, desinformación, innovación, formación, trazabilidad y sesgos. A través de un análisis fundamentado en el compendio de seis artículos, esta tesis sostiene que la IA no constituye una tecnología neutra, sino un vector cultural y político atravesado por condiciones sociales y profesionales. Partiendo de esta premisa, el trabajo se articula en torno a la pregunta central de la investigación: ¿Cómo está transformando la IA el ecosistema informativo vasco y bajo qué condiciones éticas debe gobernarse dicha transformación? Responder a este interrogante permite desgranar las tensiones entre la innovación técnica y la preservación de los valores periodísticos en un contexto de bilingüismo y proximidad territorial. La mirada propuesta combina marcos de estudios de Ciencia, Tecnología y Sociedad (CTS) con la epistemología crítica para analizar la automatización desde sus efectos concretos en el territorio. En última instancia, el trabajo concluye con dos propuestas aplicadas: un programa formativo integral y un marco normativo para un uso ético y responsable, buscando traducir la evidencia empírica en herramientas prácticas que respondan a las necesidades urgentes de un ecosistema que busca garantizar la trazabilidad, la transparencia y el valor estratégico de las tecnologías plurilingües.

1.2. Justificación de la unidad temática

La cohesión de esta tesis doctoral por compendio no reside únicamente en su objeto de estudio, sino que lo hace también en un marco analítico diseñado para analizar la IA como un fenómeno sociotécnico situado. Mientras que la literatura científica global suele abordar habitualmente la automatización desde prismas anglosajones o puramente tecnológicos, esta investigación justifica su unidad en la necesidad de observar cómo dichas transformaciones operan en un ecosistema de proximidad con una densidad mediática y una singularidad lingüística específicas. Lejos de constituir piezas independientes, estos seis estudios están interconectados, de forma que cada fase de la investigación recoge las tensiones detectadas en la anterior para así poder profundizarlas o contrastarlas desde nuevas metodologías.

Esta unidad se manifiesta en la transición lógica desde el diagnóstico macro hasta la propuesta aplicada. Los hallazgos presentados en el primer bloque del compendio — referentes a la adopción cauta y las barreras éticas del sector— no se presentan como datos aislados, sino que actúan como el cimiento que justifica la exploración posterior de dimensiones críticas. Como se ilustra en la **Figura 1**, la arquitectura de la tesis se expande de forma radial desde un núcleo central hacia seis ejes temáticos interdependientes que abarcan desde la ética y la formación hasta la sostenibilidad lingüística y la lucha contra la desinformación.

Figura 2: Estructura de la IA en los medios del País Vasco



Fuente: Elaboración propia

De este modo, la progresión orgánica de los artículos permite que la IA sea analizada no como una herramienta de software estandarizada, sino como una fuerza que

reconfigura los pilares del periodismo —veracidad, autoría y responsabilidad— bajo los condicionamientos institucionales y culturales propios del territorio. La interconexión temática se hace especialmente evidente al analizar la relación entre la estructura del mercado vasco y la innovación responsable. La coexistencia de un potente servicio público junto a grupos multimedia de gran penetración territorial justifica un análisis que no solo atienda a la eficiencia técnica, sino también a la sostenibilidad de la información en lengua minorizada. Esta dimensión lingüística actúa como un eje transversal que atraviesa la tesis, conectando la experimentación con modelos de lenguaje de gran escala (LLMs) con la necesidad de una soberanía digital que proteja la pluralidad informativa. Por ello, la unidad del compendio se refuerza al demostrar que la "cautela informativa" detectada en las redacciones locales es, en realidad, un mecanismo de defensa editorial frente a la opacidad de los sistemas globales.

Esta unidad temática se ve reforzada, finalmente, por la transferencia de los hallazgos hacia una vertiente propositiva. El uso de un marco analítico unificado en todos los artículos —basado en ejes como la trazabilidad, la ética algorítmica y la capacitación profesional— facilita que el diagnóstico de la realidad vasca se traduzca en criterios operativos para la gobernanza del sector. De este modo, la tesis no solo identifica las tensiones del ecosistema, sino que aprovecha la consistencia de sus resultados para sugerir líneas de actuación que busquen alinear la innovación tecnológica con la función social del periodismo en Euskadi.

1.3. Ejes transversales de la investigación

La selección del ecosistema mediático vasco como objeto de estudio se justifica por la relevancia del caso para el campo de la comunicación. En un escenario académico dominado por investigaciones centradas en grandes plataformas tecnológicas y sistemas mediáticos anglófonos, el análisis del caso vasco permite contrastar los discursos generales sobre la IA con una realidad local. Al situar el foco en un territorio con una lengua minorizada y una estructura de medios descentralizada, emergen tensiones estructurales que suelen quedar invisibilizadas en los diagnósticos globales. Entre estas, destaca la brecha de justicia algorítmica derivada de la disparidad en el desarrollo tecnológico de lenguas con menor presencia digital. Como se sostiene a lo largo de esta tesis, la escasez de recursos lingüísticos no es un problema técnico aislado, sino un factor que condiciona la eficacia, la equidad y la veracidad de las herramientas automatizadas en las rutinas informativas (García-Pablos et al., 2024; Salas-Pilco et al., 2022).

Este planteamiento se materializa de forma progresiva a través de los seis artículos que componen este compendio, los cuales han sido diseñados para funcionar como

un sistema interpretativo unificado. El punto de partida es el primer estudio, aceptado en la revista *Transdigital* bajo el título “La inteligencia artificial en el periodismo: impacto social, ética y transparencia en el contexto del País Vasco”. Este trabajo proporciona la base empírica al analizar el estado de la adopción tecnológica mediante una triangulación metodológica que incluye una encuesta a 501 profesionales y entrevistas en profundidad con responsables de innovación de medios de referencia como EITB, Cadena SER Euskadi, Berria, El Correo, Naiz y Deia. Los hallazgos de este diagnóstico inicial revelan que la IA no se integra como un proceso uniforme, sino como una negociación constante entre la eficiencia técnica y la identidad editorial. Este artículo permite articular las preguntas transversales que guían el resto del compendio: desde la gobernanza institucional y la transparencia en el uso de algoritmos hasta la urgencia de una alfabetización crítica que permita a las redacciones mantener su autonomía frente a modelos entrenados mayoritariamente bajo lógicas lingüísticas y culturales ajenas al contexto vasco.

A partir de este marco general, la investigación se adentra en la dimensión técnica y territorial a través del segundo artículo, “La inteligencia artificial generativa y la información local y regional: análisis de cinco modelos de lenguaje”, publicado en *Hipertext.net*. Este estudio asume un enfoque experimental para evaluar el rendimiento de ChatGPT, Claude, Gemini, Copilot y Perplexity ante consultas sobre la actualidad vasca reciente. Al analizar casos mediáticos de alta sensibilidad —como la gestión sanitaria o procesos judiciales de relevancia local—, se evidencia una limitación estructural en los Modelos de Lenguaje de Gran Escala (LLMs): la pérdida de trazabilidad y la degradación del pluralismo informativo en escalas regionales. Los resultados demuestran que, cuando la IA opera fuera de los grandes centros de datos globales, tiende a simplificar la complejidad institucional y a invisibilizar fuentes locales, lo que refuerza la necesidad de implementar marcos de evaluación que incorporen dimensiones contextuales y no solo métricas de rendimiento sintáctico (Ermolova et al., 2024).

La dimensión pedagógica y organizacional se aborda en el tercer trabajo, publicado en *Frontiers in Communication* con el título “Artificial intelligence training in media: addressing technical and ethical challenges for journalists and media professionals”. En él se constata un déficit estructural en la capacitación: la mayoría de los periodistas vascos depende del autoaprendizaje, lo que genera una vulnerabilidad ética y operativa en las redacciones. Se sostiene que este vacío limita la capacidad de los profesionales para supervisar de manera crítica las lógicas algorítmicas, ampliando la brecha entre la innovación técnica y la gobernanza editorial. El artículo propone un modelo de competencias que trasciende lo instrumental para centrarse en la toma de decisiones fundamentadas, reforzando la idea de que la falta de formación es, en última instancia, una barrera para la soberanía profesional en la era de la automatización (Sonni et al., 2024).

Tras delimitar estas condiciones materiales, el compendio profundiza en la apropiación sociotécnica a través del cuarto artículo, “La inteligencia artificial como

motor de innovación en los medios: apropiación técnica y cautela informativa”, publicado en *Estudios sobre el Mensaje Periodístico*. Este estudio introduce el concepto de “cautela informativa”, entendido como una estrategia deliberada de los medios vascos para integrar la IA de forma progresiva y controlada. El análisis de las prácticas en redacciones como EITB permite concluir que la IA no está sustituyendo el trabajo humano, sino que se está insertando en los procesos de producción bajo una supervisión editorial estricta. Esta perspectiva se alinea con las teorías que ven la tecnología como un proceso moldeado por los valores del oficio, donde la resistencia o la lentitud en la adopción no es signo de atraso, sino de una responsabilidad editorial que busca proteger la credibilidad ante la opacidad algorítmica (Pinto & Barbosa, 2024).

El quinto artículo, “Kazetarietzako tresnak adimen artifizialak indartutako desinformazioaren aurka”, publicado en *Mediatika*, examina las herramientas y estrategias de verificación que emplean los profesionales vascos para identificar manipulaciones automatizadas y contenidos sintéticos. Este estudio dialoga estrechamente con la evaluación técnica de los LLMs realizada anteriormente: si aquel detectaba los fallos en la generación de información, este explora la capacidad del campo profesional para responder a ellos. Se concluye que la lucha contra la desinformación en un entorno multilingüe requiere no solo herramientas avanzadas, sino infraestructuras colaborativas entre medios, universidades y reguladores.

Finalmente, el sexto artículo, “Analysis on disinformation and gender bias generated by artificial intelligence”, enviado a *Interconectando Saberes*, aporta la necesaria capa de justicia epistémica. El estudio analiza de manera interseccional cómo los sistemas algorítmicos pueden reproducir y amplificar sesgos de género en la cobertura informativa automatizada. Al integrar conceptos como la “opacidad sistémica”, el trabajo subraya que la IA no puede analizarse al margen de sus efectos sociales y culturales. Su incorporación al compendio refuerza la vocación transformadora de la tesis, demostrando que la innovación tecnológica solo es legítima si se acompaña de un compromiso innegociable con la equidad, la inclusión y la representación justa de la diversidad social.

A continuación, se presenta una síntesis de la producción científica que sustenta la tesis, articulada no como una suma de partes, sino como una estructura de conocimiento interconectada. Esta disposición permite a su vez poder verificar la solidez de la triangulación metodológica empleada y la coherencia del itinerario investigador seguido.

La siguiente tabla compacta la información relativa a las publicaciones, ofreciendo una panorámica de su impacto editorial y su cronología.

Tabla 3: Artículos publicados de la tesis por compendio:

Nº	Título del Artículo	Revista	Idioma	Envío	Aceptación	Publicación
1	"La IA en el periodismo: impacto, ética y transparencia..."	<i>Transdigital</i>	Cast.	05/07/25	31/07/25	—
2	"IA generativa y medios locales: sesgo territorial..."	<i>Hipertext.net</i>	Cast.	15/07/25	13/10/25	25/11/25
3	"Artificial intelligence training in media..."	<i>Frontiers in Comm.</i>	Inglés	03/12/24	20/05/25	09/06/25
4	"La IA como motor de innovación: apropiación y cautela..."	<i>Estud. Mensaje Per.</i>	Cast.	25/05/25	18/06/25	10/09/25
5	"Kazetarientzako tresnak desinformazioaren aurka"	<i>Mediatika</i>	Eusk.	18/12/24	—	—
6	"Analysis on disinformation and gender bias..."	<i>Interc. Saberes</i>	Inglés	04/11/25	—	—

Fuente: Elaboración propia

Más allá de su cronología, la verdadera fortaleza de esta tesis reside en el ensamblaje de sus dimensiones analíticas. Cada artículo ha sido concebido para cubrir un vacío específico, de modo que la resolución de los interrogantes planteados en un estudio se convierte en la base empírica del siguiente. Para analizar esta estructura, la Tabla 2 detalla la correspondencia entre los ejes temáticos nucleares y las publicaciones, evidenciando cómo la triangulación de métodos —encuestas, entrevistas, experimentación controlada y análisis documental— garantiza una perspectiva plural sobre el fenómeno.

Tabla 4: Correspondencia entre los principales ejes temáticos de la tesis doctoral y las publicaciones del compendio

Ejes temáticos principales	P 1	P 2	P 3	P 4	P 5	P 6
IA y transformación del periodismo (adopción, producción)	X			X		
Ética, transparencia y responsabilidad en el uso de IA	X		X	X		X
Gobernanza y regulación de la IA en medios	X		X	X	X	
Desinformación y verificación automatizada		X			X	X
Sostenibilidad lingüística (lengua vasca y diversidad cultural)	X	X		X	X	
Formación y alfabetización en IA	X		X			
Sesgos algorítmicos y equidad (p. ej., género)						X

Fuente: Elaboración propia

En definitiva, la tesis no se limita a una descripción funcional de herramientas técnicas, sino que analiza las profundas consecuencias sociales y culturales de la automatización informativa en un territorio con identidad propia. Al integrar las dimensiones de desinformación, formación profesional y justicia lingüística, el compendio ofrece una interpretación integral de los retos éticos de la era algorítmica. Esta visión se alinea con las corrientes académicas actuales que reivindican una gobernanza responsable (Yeung & Dodds, 2024; Gutiérrez-Caneda et al., 2024), presentando una aportación original que demuestra que la transformación digital del periodismo solo puede comprenderse plenamente cuando se atiende a la especificidad de los contextos locales y multilingües.

1.4. Estructura general de la tesis

La estructura de esta tesis doctoral se ha organizado para cumplir con la normativa de la Universidad del País Vasco (UPV/EHU) relativa a la modalidad de compendio de publicaciones, garantizando al mismo tiempo una secuencia lógica que otorgue cohesión al proyecto investigador. El diseño no se concibe como una mera agregación de seis artículos científicos independientes, sino como una arquitectura integrada

donde cada pieza contribuye a un discurso global sobre la transformación digital del periodismo vasco.

La arquitectura de la tesis se articula en torno a tres grandes bloques estructurales que guían al lector desde el planteamiento conceptual hasta la propuesta de transferencia de conocimiento, tal como se detalla a continuación:

Tabla 5: Organización detallada de la tesis por bloques y contenidos

Bloque Estructural	Capítulos e Hitos	Propósito y Contenido Analítico
I. Marco de Referencia y Diseño	1. Introducción 2. Objetivos 3. Hipótesis 4. Marco Teórico 5. Metodología	Establece la unidad temática del compendio. Define los pilares teóricos (automatización, gobernanza algorítmica y soberanía tecnológica) y justifica la relevancia del caso vasco. Describe el diseño metodológico mixto (encuestas, entrevistas y experimentos) que sostiene la investigación.
II. Resultados e Integración Crítica	6. RESULTADOS 7. Discusión	Constituye el núcleo analítico. El capítulo de Resultados sintetiza los hallazgos de los seis artículos publicados, mientras que la Discusión eleva esos datos al debate científico internacional, contrastándolos con las dinámicas globales de la IA.
III. Síntesis Final y Transferencia	8. Conclusiones	Recapitula las aportaciones originales y las implicaciones del estudio. Los anexos integran el cuerpo empírico (los seis artículos) y los productos

	9. Limitaciones 10. Futuras líneas 11. Anexos	de transferencia: un programa formativo y un marco normativo ético para las redacciones.
--	---	--

Fuente: Elaboración propia.

Esta organización permite que la tesis presente una estructura lógica y progresiva. En el primer bloque, se establece el marco teórico y se delimitan los objetivos de investigación, proporcionando el contexto necesario para entender la singularidad del ecosistema bilingüe vasco. Es aquí donde se plantean las hipótesis que serán validadas o matizadas en los capítulos posteriores.

El segundo bloque representa el aporte fundamental del trabajo. A diferencia de una tesis tradicional, el capítulo de "Discusión" permite la triangulación de los resultados de los artículos. En esta sección se analiza de forma transversal cómo la IA impacta en la autonomía editorial, la calidad informativa y la representación de la diversidad lingüística, permitiendo extraer patrones que superan la suma de las partes individuales.

Finalmente, el tercer bloque traslada el diagnóstico a la acción. Las conclusiones no solo cierran los interrogantes planteados, sino que se materializan en los anexos mediante propuestas aplicadas. Estas herramientas (el programa de formación y el marco normativo) se han diseñado para responder a las carencias específicas detectadas en los artículos, como el déficit formativo de los profesionales y la ausencia de protocolos éticos en los medios.

En definitiva, la estructura propuesta asegura que la investigación no sea solo un análisis descriptivo, sino un proyecto con vocación transformadora. Al vincular estrechamente los resultados empíricos con propuestas de intervención profesional, el trabajo se posiciona como una contribución situada que ofrece soluciones concretas para la integración ética de la IA en el periodismo.

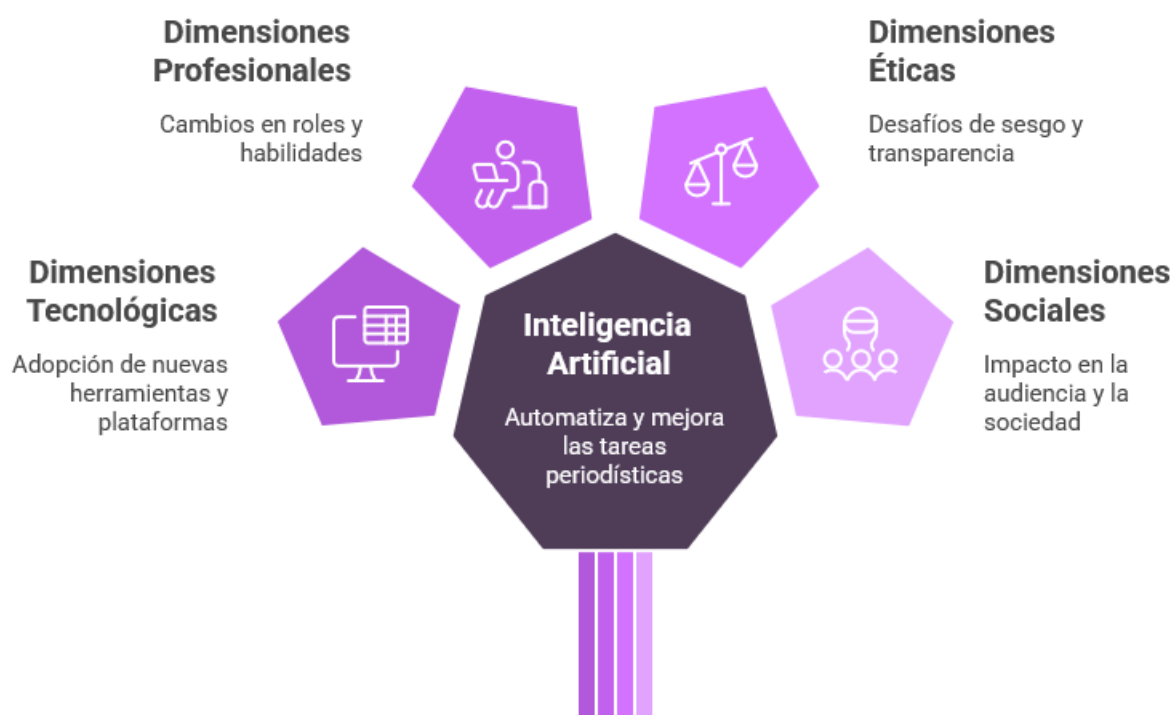
2. Objetivos

A partir de las hipótesis planteadas, se definió un objetivo general que guía la investigación, desglosado en seis objetivos específicos (OE1–OE6) vinculados a sus correspondientes hipótesis de investigación secundarias (H1–H6). A continuación, se presenta el objetivo general y cada objetivo específico, conectándolos explícitamente con las hipótesis a las que responden:

2.1. Objetivo general

El objetivo general de esta tesis es analizar el impacto de la IA en el ejercicio del periodismo en el contexto del País Vasco, atendiendo a sus dimensiones tecnológicas, profesionales, éticas y sociales. Para ello se adopta un enfoque empírico basado en estudios de caso, encuestas, entrevistas y análisis documental, con el fin de contribuir a una comprensión situada y profunda de sus oportunidades, desafíos y riesgos, especialmente en entornos sociolingüísticos diversos y escenarios de transición digital acelerada.

Figura 6: Objetivo general de la investigación



Fuente: Elaboración propia

2.2. Objetivos específicos

OE1: Examinar los efectos de la implementación de sistemas de IA generativa en las dinámicas laborales, organizativas y editoriales de medios tradicionales del País Vasco, a través del análisis de casos. Se corresponde con la Hipótesis secundaria 1, que plantea a su vez que la incorporación de tecnologías generativas de IA en medios vascos está modificando significativamente las rutinas profesionales del periodismo, alterando de este modo la distribución de tareas, los roles de los redactores y los criterios de edición de contenidos.

OE2: Analizar las percepciones, actitudes y niveles de alfabetización tecnológica en IA entre periodistas del País Vasco, considerando factores como la formación profesional, el entorno mediático y la existencia (o ausencia) de marcos normativos claros. Este objetivo responde a la Hipótesis secundaria 2, la cual sostiene por su lado que las percepciones de los profesionales sobre la IA están mediadas por la formación previa, el tipo de medio, el acceso a recursos tecnológicos y la claridad de los marcos normativos, generando así una tensión entre la apertura a la innovación y la preocupación ética.

OE3: Evaluar la eficacia, adecuación contextual y barreras de uso de las herramientas automatizadas de generación de contenido basadas en IA, especialmente en entornos multilingües como el euskera, y su integración en las rutinas de trabajo periodístico. Se formula en línea con la Hipótesis secundaria 3, que afirma que las herramientas de generación automatizada basadas en IA presentan limitaciones de precisión y adaptabilidad cuando se aplican en contextos multilingües como el vasco, lo que compromete su eficacia en la detección y neutralización de la desinformación.

OE4: Identificar y analizar los sesgos de género presentes en los sistemas algorítmicos aplicados a la generación y distribución de contenidos mediáticos, así como sus implicaciones en la representación y visibilidad de las mujeres en la esfera informativa. Se vincula con la Hipótesis secundaria 4, que advierte que la automatización de procesos informativos mediante IA puede amplificar sesgos de género y reforzar desigualdades preexistentes si no se somete a mecanismos rigurosos de auditoría ética, revisión editorial y control profesional.

OE5: Examinar los mecanismos —o la falta de ellos— de transparencia, gobernanza y rendición de cuentas en el uso de IA por parte de medios de comunicación, y su influencia en la ética periodística, la autonomía profesional y la confianza de la ciudadanía. Corresponde a la Hipótesis secundaria 5, que postula que la ausencia de mecanismos sólidos de transparencia y rendición de cuentas en el uso de IA por los medios debilita la confianza pública y compromete principios fundamentales del periodismo democrático.

OE6: Contextualizar los hallazgos empíricos de los casos vascos en el marco más amplio de las políticas y marcos normativos europeos sobre IA, periodismo y

democracia, identificando tensiones, brechas regulatorias y posibilidades de desarrollo ético. Este objetivo se relaciona con la Hipótesis secundaria 6, que sostiene que la integración acelerada de IA en los medios exige nuevas formas de gobernanza adaptadas a la especificidad sociolingüística y cultural de contextos periféricos como el vasco, donde las dinámicas de poder tecnológico global no siempre coinciden con las prioridades comunicativas locales.

Figura 7: Objetivos específicos de la investigación



Fuente: Elaboración propia

2.3. Correspondencia entre objetivos y artículos

La arquitectura de esta tesis se fundamenta en una correlación directa entre los interrogantes teóricos y la producción científica resultante. Cada publicación que integra este compendio ha sido diseñada para dar respuesta a uno o varios de los

objetivos específicos, garantizando así que la hipótesis general sea contrastada desde múltiples dimensiones: la profesional, la técnica, la ética y la social.

A diferencia de una estructura lineal, esta tesis apuesta por una triangulación de resultados que permite una comprensión holística del fenómeno:

- Impacto en la cultura profesional (OE1 y OE4): El análisis de la "cautela informativa" y la apropiación técnica (Artículo 1) permite comprender que la IA no es un agente neutro, sino que por el contrario su integración está mediada por los valores del oficio. Este enfoque se completa con la mirada crítica sobre los sesgos de género (Artículo 6), demostrando que la innovación tecnológica solo alcanza su plenitud si corrige las desigualdades estructurales en la representación mediática.
- Competencias y marcos regulatorios (OE2 y OE6): La investigación profundiza en la brecha formativa (Artículo 3), conectando el déficit de capacitación detectado en el territorio con las exigencias de los marcos normativos europeos. Se evidencia que la alfabetización tecnológica no es solo una competencia técnica, sino un requisito para el ejercicio ético del periodismo.
- Desafíos lingüísticos y transparencia (OE3 y OE5): Al evaluar el rendimiento de los modelos de lenguaje en euskera (Artículo 2) y las herramientas de verificación (Artículo 5), la tesis aporta evidencia sobre la soberanía tecnológica. Se demuestra que la opacidad de los algoritmos y su limitado desempeño en lenguas minorizadas son barreras directas para la rendición de cuentas (Artículo 4) que el periodismo debe a su audiencia.

Tabla 8. Correspondencia entre objetivos específicos (OE) y artículos del compendio

Objetivos específicos (OE)	1	2	3	4	5	6
OE1. Efectos de la IA generativa en dinámicas laborales/editoriales (estudios de caso)	✓					
OE2. Percepciones, actitudes y alfabetización en IA entre periodistas		✓	✓		✓	

OE3. Eficacia y barreras de la generación automatizada de contenidos (contexto multilingüe)		✓				
OE4. Sesgos de género en sistemas algorítmicos aplicados a contenidos						✓
OE5. Transparencia, gobernanza y rendición de cuentas en el uso de IA				✓		
OE6. Contextualización frente a marcos normativos y políticas europeas			✓			

Fuente: Elaboración propia

Leyenda de artículos

1. *La IA como motor de innovación en los medios: apropiación técnica y cautela informativa* (EMP).
2. *La IA generativa y la información local y regional: análisis de cinco modelos de lenguaje* (Hiptertext.net).
3. *Artificial intelligence training in media: addressing technical and ethical challenges for journalists and media professionals* (Frontiers in Communication).
4. *La IA en el periodismo: impacto social, ética y transparencia en el contexto del País Vasco* (Transdigital).
5. *Kazetarienezako tresnak adimen artifizialak indartutako desinformazioaren aurka* (Mediatika).
6. *Analysis on disinformation and gender bias generated by artificial intelligence* (Interconectando Saberes).

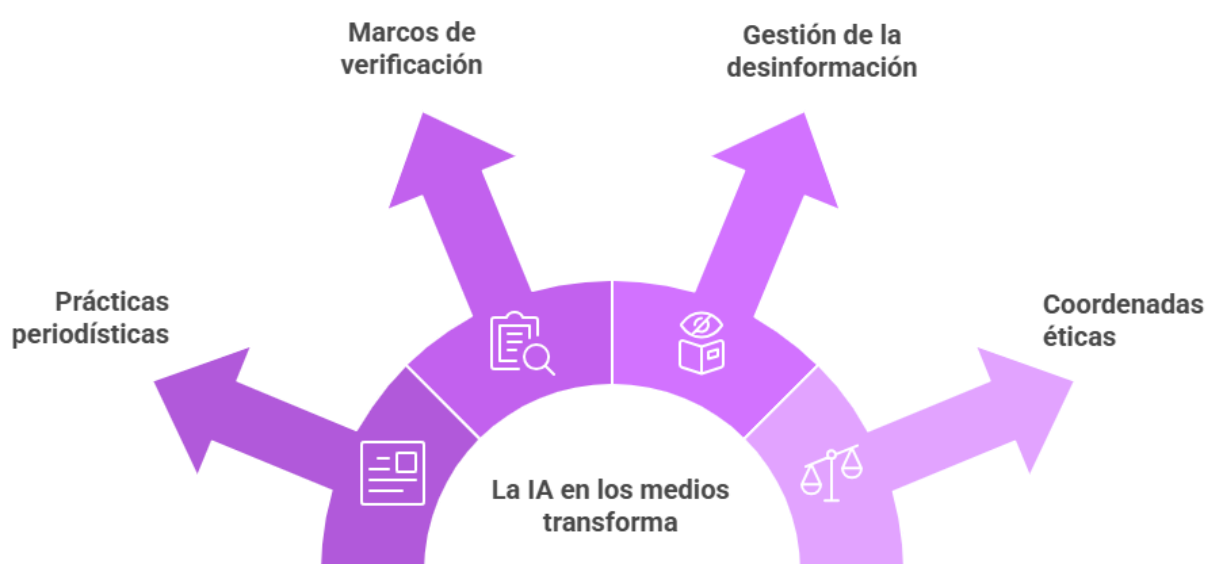
Esta interdependencia asegura que la tesis no solo identifique problemas, sino que los analice mediante una base empírica diversificada. La coherencia del conjunto se garantiza al demostrar que cada artículo, desde su especificidad metodológica —ya sea el rigor estadístico de la encuesta o la profundidad cualitativa del estudio de caso—, contribuye de manera unívoca a responder a la pregunta central de esta investigación: *¿Cómo está transformando la IA el ecosistema informativo vasco y bajo qué condiciones éticas debe gobernarse dicha transformación?*

3. Hipótesis de investigación

3.1. Hipótesis principal

La progresiva incorporación de sistemas de IA en los medios de comunicación del País Vasco está transformando de manera estructural las prácticas periodísticas, los marcos de verificación, la gestión de la desinformación y las coordenadas éticas del trabajo informativo, generando nuevas tensiones profesionales, riesgos socio-técnicos y desafíos de gobernanza, especialmente en contextos regionales multilingües con alfabetización digital desigual.

Figura 8: Hipótesis principal de la investigación



Fuente: Elaboración propia

3.2. Hipótesis secundarias

Hipótesis secundaria 1:

La implementación de tecnologías generativas de IA en medios del País Vasco está modificando significativamente las rutinas de los profesionales del periodismo, alterando la distribución de las tareas, el rol de los redactores y los criterios de edición de los contenidos.

Hipótesis secundaria 2:

Las percepciones de los profesionales sobre la IA están mediadas por factores como la formación previa, el tipo de medio, el acceso a recursos tecnológicos y la

claridad de los marcos normativos, lo que genera una tensión entre apertura a la innovación y preocupación ética.

Hipótesis secundaria 3:

La IA generativa presenta limitaciones de precisión, trazabilidad y adaptabilidad en contextos multilingües como el vasco, lo que compromete tanto la calidad de los contenidos producidos como su integración en rutinas periodísticas y en estrategias de gestión de la desinformación.

Hipótesis secundaria 4:

La automatización de procesos informativos mediante IA tiende a amplificar sesgos de género y reforzar desigualdades preexistentes si no se somete a mecanismos rigurosos de auditoría ética, revisión editorial y control profesional.

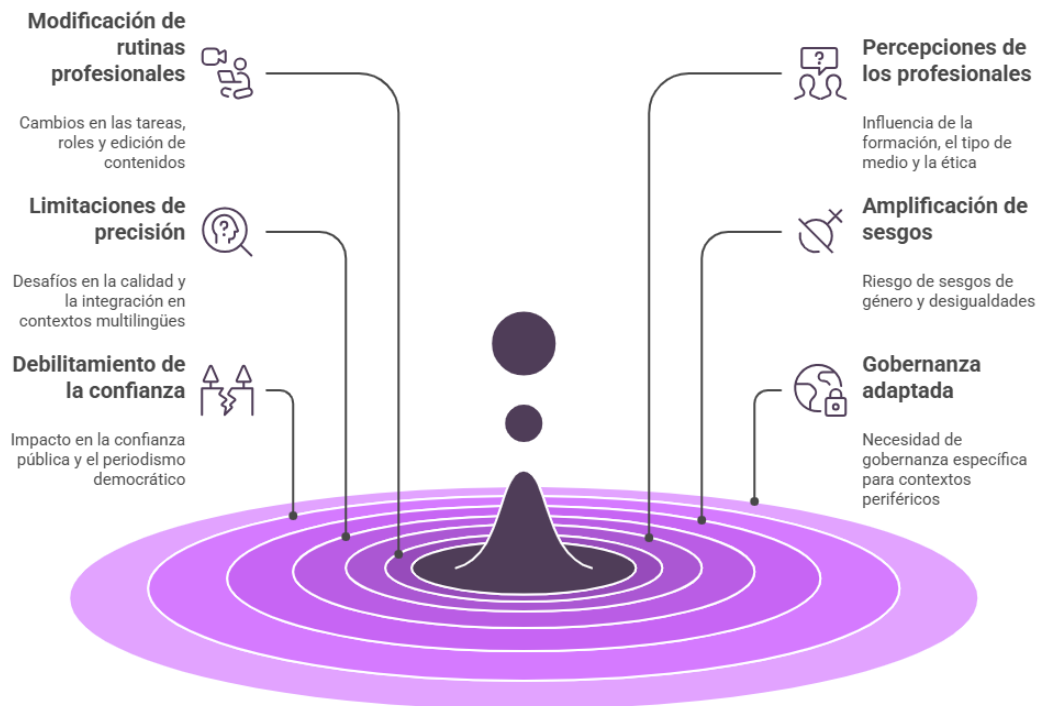
Hipótesis secundaria 5:

La ausencia de mecanismos sólidos de transparencia y rendición de cuentas en el uso de IA por parte de medios de comunicación debilita la confianza pública y compromete principios fundamentales del periodismo democrático.

Hipótesis secundaria 6:

La integración acelerada de IA en los medios exige nuevas formas de gobernanza, adaptadas a la especificidad sociolingüística y cultural de contextos periféricos, como el vasco, donde las dinámicas de poder tecnológico global no siempre coinciden con las prioridades comunicativas locales.

Figura 9: Hipótesis secundarias de la investigación



Fuente: Elaboración propia

4. Marco teórico

4.1. Principios y aproximaciones teóricas

Esta investigación se sitúa en la intersección crítica entre la comunicación, la inteligencia IA y la alfabetización mediática, partiendo de una premisa que atraviesa cada fase de este trabajo: la IA se define no solo como una herramienta técnica, sino como un fenómeno sociotécnico que reconfigura las estructuras fundamentales del periodismo contemporáneo. Para comprender el alcance de esta transformación, resulta imperativo sistematizar el mapa de tensiones que esta tecnología proyecta sobre la esfera pública. Esta transición no depende exclusivamente del desarrollo de software, sino de la capacidad crítica de los sujetos profesionales para interpretar y gobernar estos sistemas. No se trata únicamente de la adopción de herramientas inéditas, sino de una metamorfosis que afecta a la raíz de la profesión: desde la atribución de responsabilidades y la dilución de la autoría, pasando por la gobernanza algorítmica de los procesos editoriales, hasta el cuestionamiento de la neutralidad tecnológica ante la reproducción de sesgos que marginan a lenguas y culturas con menor densidad digital.

El primer desafío de orden ontológico que emerge es la redefinición de la autoría informativa. Montal y Reich (2016) han analizado cómo la automatización produce un desplazamiento de la firma periodística hacia los sistemas técnicos, lo que deriva en una opacidad creciente al intentar delimitar quién debe rendir cuentas ante posibles errores o imprecisiones. Esta transición hacia un periodismo de naturaleza algorítmica no constituye un cambio de soporte, sino un movimiento que reordena las relaciones de poder. En este sentido, Dörr y Hollnbuchner (2016) y Helberger et al. (2022) sostienen que nos hallamos en una fase de renegociación constante de los roles entre medios, plataformas y audiencias, a menudo bajo lógicas impuestas por las grandes infraestructuras globales de datos que condicionan la autonomía del relato. Esta renegociación de roles se manifiesta en la práctica como una resistencia fundamentada en la preservación del capital simbólico del periodista. La relevancia de este debate ontológico se confirma al observar que la opacidad de los sistemas no es solo una preocupación teórica, sino una barrera operativa real: los datos del Artículo 1 de esta tesis muestran que el 63,91% de los profesionales identifica los dilemas éticos y la falta de transparencia algorítmica como el principal freno para la adopción de la IA, lo que subraya la vigencia de la crítica de Dörr y Hollnbuchner (2016) sobre la soberanía editorial.

Precisamente para mitigar esta desconfianza y preservar el control editorial en este ecosistema, la literatura científica subraya la necesidad de implementar mecanismos de *accountability* o rendición de cuentas algorítmica. Diakopoulos (2015) y Fairfield y Shtein (2014) definen este principio no solo como la aplicación de la IA, sino como la capacidad institucional de explicar y auditar las decisiones automatizadas que subyacen a la jerarquización informativa. Esta exigencia de transparencia resulta vital al afectar directamente a la veracidad, pilar de la función social del periodismo. Al respecto, Ekström y Westlund (2019) y Pavlik (2014) advierten sobre la urgencia de

actualizar los marcos de verificación en entornos de *big data*, mientras que Godler et al. (2020) analizan el peso determinante que las plataformas digitales han adquirido en la construcción social de la realidad. Bajo esta perspectiva, la propuesta de Guzman y Lewis (2024) adquiere una relevancia estratégica: la IA opera como un agente que co-participa en la creación del discurso público, transformando la dinámica de producción de noticias y exigiendo niveles de alfabetización que impidan la delegación del criterio profesional en la máquina.

Este análisis profundo obliga a trascender la visión puramente instrumental de la tecnología para adoptar una perspectiva de corte crítico. El análisis bibliométrico de Ioscote et al. (2024) confirma que, especialmente desde 2018, la investigación académica ha experimentado un giro terminológico y conceptual: el interés se ha desplazado de la eficiencia técnica hacia el impacto de la IA en la salud democrática de las sociedades. Trabajos como los de Bayer (2024) o Peña-Fernández et al. (2023) han resultado determinantes al integrar variables de género, lengua y justicia de datos en la discusión comunicativa. Es en este punto donde las aportaciones de Lewis et al. (2019) y Latzer y Just (2020) permiten comprender que la tecnología no es neutral. Por el contrario, estudios como los de Amponsah & Atianashie (2024) y Bailon (2024) demuestran que los algoritmos tienden a heredar y amplificar las desigualdades estructurales presentes en el entorno analógico. Cuéllar Rodríguez (2023) y Mohamed et al. (2020) profundizan en esta tesis al recordar que la IA se despliega siempre dentro de coordenadas de poder económico y geopolítico muy definidas, lo que obliga al periodista a desarrollar una competencia crítica superior para salvaguardar su independencia.

Dichas dinámicas de poder tienen consecuencias tangibles en lo relativo a la justicia de datos y la diversidad cultural. Ermolova et al. (2024) han aportado evidencia sobre cómo los modelos de IA entrenados en lenguas mayoritarias suelen omitir o interpretar de forma deficiente las realidades de territorios con lenguas minorizadas. Este fenómeno genera lo que Mohamed et al. (2020) y Jandrić (2019) definen como una vulnerabilidad epistémica: si las herramientas de producción informativa no codifican correctamente el contexto o el idioma local, se debilita la capacidad de las comunidades para articular sus propias narrativas. La urgencia de este enfoque sobre la justicia de datos se agrava ante la 'paradoja de la adopción' global. Mientras que informes de Google Cloud (2024) señalan que el 74% de las grandes corporaciones ya perciben un retorno de inversión directo de la IA, el periodismo en lenguas minorizadas enfrenta una 'brecha de utilidad' debido a la escasez de datasets específicos. Esta asimetría, analizada experimentalmente en el Artículo 2, demuestra que el rendimiento de los modelos de lenguaje (LLMs) decae sistemáticamente al procesar contextos locales vascos, validando la tesis de Mohamed et al. (2020) sobre la vulnerabilidad epistémica en las periferias digitales. En este marco, la decisión de las redacciones de proceder con una "cautela informativa" no debe entenderse como un rechazo a la innovación, sino como una práctica de resistencia ética orientada a proteger la autonomía editorial frente a sistemas inherentemente opacos.

La veracidad de este andamiaje teórico remite necesariamente a la importancia del factor humano y su capacitación. La tecnología solo adquiere una dimensión ética si los profesionales poseen el bagaje analítico necesario para gestionarla desde la sospecha crítica. Salgado (2022), Connor y Cali (2024) y Lopezosa et al. (2023) coinciden en que la alfabetización mediática contemporánea debe superar el aprendizaje operativo de herramientas para centrarse en una competencia profunda que permita detectar sesgos, supervisar la producción automatizada y mantener la integridad del relato periodístico. Sin esta base formativa, el periodismo corre el riesgo de erosionar su función de vigilancia social. Por ello, este marco teórico sostiene que la integración de la IA en las redacciones es un proceso sociotécnico de alta complejidad que requiere una gobernanza proactiva y una ciudadanía profesional formada, capaz de garantizar que la innovación tecnológica no comprometa los valores esenciales del periodismo ni la pluralidad lingüística de la esfera pública.

Finalmente, la construcción de este marco ético responde a una demanda social creciente. Según el Reuters Institute (2024), el público global manifiesta una desconfianza mayoritaria (52%) hacia las noticias generadas principalmente por IA, lo que dota de una dimensión práctica a la teoría de la gobernanza proactiva: la transparencia algorítmica no es solo un ideal académico, sino el único camino para restaurar la confianza de una audiencia que, ante la proliferación de contenidos sintéticos, exige una supervisión humana verificable. De este modo, los principios aquí desgranados conforman un vocabulario analítico capaz de interpretar la realidad del periodismo en entornos de transición digital acelerada. Al conectar la ética del dato con la soberanía lingüística y la formación crítica, se establece una base sólida para analizar cómo los medios de proximidad enfrentan la hegemonía de los modelos globales. Este marco teórico, por tanto, no solo diagnostica la situación actual, sino que fundamenta la necesidad de una gobernanza que priorice la justicia informativa y el derecho de las sociedades a recibir una información veraz, transparente y representativa de su propia diversidad, independientemente de la sofisticación de los sistemas que intervengan en su producción.

La complejidad de los fenómenos descritos —desde la dilución de la autoría hasta la justicia de datos en contextos multilingües— exige que los conceptos nucleares de esta tesis no solo se enuncien, sino que se delimiten bajo un marco operativo claro. Por ello, una vez establecidas las corrientes de pensamiento que sustentan la investigación, el siguiente apartado se dedica a la definición de los conceptos y definiciones (2.1.1) que constituirán el vocabulario analítico de este trabajo.

4.2. Conceptos y definiciones

Para analizar la integración de la IA en el ecosistema mediático del País Vasco, resulta necesario configurar un repertorio conceptual que funcione como una lente crítica sobre la realidad bilingüe y estructuralmente diversa de Euskadi. Este planteamiento nos aleja de una simple lista de etiquetas abstractas para ofrecernos una caja de herramientas interpretativa, útil para desgranar cómo la lógica algorítmica colisiona o se ensambla con la función social del periodismo de proximidad. Con este fin, el apartado se organiza a través de cuatro grandes dimensiones que estructuran el discurso: la esfera operativa y productiva; la vertiente de integridad y riesgos; el plano de la gobernanza y la ética; y, finalmente, el eje de la capacitación humana.

Al observar la esfera de la producción, el concepto de periodismo automatizado se sitúa en el centro del debate. Entendido como la generación de contenidos mediante algoritmos sin intervención humana directa (Danzon-Chambaud, 2021; Thurman et al., 2017), este fenómeno plantea interrogantes de calado sobre la calidad contextual y la autonomía del profesional. La literatura advierte que delegar tareas en sistemas técnicos puede erosionar la autoría y los estándares tradicionales del oficio (Monti, 2019; Montal & Reich, 2017; Danzon-Chambaud & Cornia, 2021). En un entorno como el vasco, este concepto ayuda a identificar el hecho de que la automatización no avanza como una fuerza sustitutiva, sino como un apoyo en tareas periféricas —etiquetado o borradores—, condicionado severamente por la falta de herramientas robustas en euskera, lo que obliga a mantener un modelo de innovación supervisada antes que delegada.

Ligada estrechamente a la producción se encuentra la personalización informativa. Aunque esta técnica busca mejorar la experiencia del usuario mediante la adaptación de la oferta, la literatura académica alerta sobre riesgos sistémicos relacionados con la privacidad, la fragmentación del pluralismo y la creación de cámaras de eco (Du, 2023; Fu, 2023). Aquí, la noción de burbuja de filtro (Bodó, 2019) adquiere una relevancia estratégica: en un territorio con una fragmentación lingüística natural, el uso intensivo de datos para la recomendación de contenidos podría amplificar dinámicas de aislamiento cognitivo, justificando un análisis preventivo antes incluso de que estas tecnologías se desplieguen de forma masiva en los medios locales.

En cuanto a la integridad del espacio público, el sesgo algorítmico emerge como una categoría crítica fundamental. Se refiere a las distorsiones en la selección o clasificación de contenidos que derivan de datos de entrenamiento sesgados o de las lógicas de diseño (Tsekhmeistruk, 2024; Carlson, 2017). Para mitigar este riesgo, la literatura propone la transparencia algorítmica, concebida no como una apertura técnica del código fuente, sino como una práctica comunicativa que rinda cuentas sobre los límites y efectos de estos sistemas (Diakopoulos & Koliska, 2017; Bitzer et al., 2022; Karlsson, 2021). Este concepto resulta vital para entender las desigualdades epistémicas que sufren las lenguas minorizadas, donde la baja

trazabilidad de los modelos generativos puede invisibilizar la realidad regional frente a la hegemonía de los grandes relatos globales.

Esta preocupación por la veracidad se extiende a la sofisticación de la desinformación contemporánea. El uso de la IA ha potenciado técnicas retóricas de saturación y deslegitimación como el *Galope de Gish*, el *cherry-picking*, el *DARVO* o la *Defensa Chewbacca* (Makeleni et al., 2023). Estos métodos buscan desbordar la capacidad analítica del receptor y del propio periodista. A nivel estratégico, fenómenos como la Manipulación Informativa Extranjera (FIMI) y técnicas de infiltración como el *ataque mariposa* (Porlezza, 2023; Schneider et al., 2023) ilustran cómo actores externos pueden desestabilizar comunidades digitales. En un contexto como el vasco, expuesto a infraestructuras globales pero con una escala demográfica pequeña, comprender estas tácticas se vuelve esencial para interpretar la presión cognitiva y profesional que sufren las redacciones.

Desde la perspectiva de la gobernanza, la responsabilidad editorial se redefine en escenarios híbridos donde la frontera entre lo humano y lo automático se vuelve porosa. Helberger et al. (2022) y Bayer (2024) subrayan que esta hibridación exige mecanismos formales de trazabilidad y una clarificación tajante de la autoría. Esto adquiere un matiz crítico ante la proliferación de *deepfakes*, la clonación de voz o la *impersonation*, fenómenos que amenazan la confianza ciudadana si no existen protocolos internos de supervisión. Para abordar estas asimetrías, el enfoque de justicia de datos proporciona un marco ético superior, integrando principios de equidad e inclusividad (Showkat, 2022; Christin, 2017; Taylor, 2017). En esta tesis, la justicia de datos permite explicar cómo la escasez de recursos tecnológicos en euskera no es solo un problema técnico, sino que también es una brecha de representación que limita la participación democrática. Por ello, los marcos éticos aplicados al periodismo deben combinar principios deontológicos clásicos con enfoques basados en valores (Kim, 2019; Prem, 2023), traducándose en herramientas operativas como la auditoría algorítmica y la estandarización de prácticas (Dierickx et al., 2024; Forja-Pena et al., 2024).

Finalmente, ninguna de estas dimensiones puede gestionarse sin situar la alfabetización en IA como el pilar que sostiene el sistema. Definida como la capacidad de comprender, evaluar y emplear críticamente los sistemas algorítmicos (Ng et al., 2021), esta alfabetización incorpora una dimensión ética y organizacional ineludible (Deuze & Beckett, 2022; Dinçer, 2024). En un ecosistema bilingüe con una brecha formativa latente, la alfabetización deja de ser una opción individual para convertirse en una necesidad estructural. El hecho de que el autoaprendizaje sea la vía predominante en las redacciones vascas solo refuerza la urgencia de establecer marcos formativos estables que permitan a los profesionales detectar sesgos y ejercer una vigilancia crítica sobre la máquina.

Todo este repertorio conceptual conforma la base analítica de la tesis. Su articulación demuestra que el periodismo no es un receptor pasivo de la tecnología, sino un campo

en disputa donde se negocian nuevas formas de agencia y producción de conocimiento. De este modo, el mapa de conceptos propuesto permite interpretar la transición digital en el País Vasco como un proceso sociotécnico donde la ética, la lengua y la capacitación humana son los verdaderos ejes que definirán el futuro de la información veraz en la era de los algoritmos.

4.3. Estado de la cuestión

El debate académico contemporáneo sobre la inteligencia IA aplicada al ámbito informativo atraviesa una fase de profunda redefinición epistemológica. Lo que hace apenas una década se analizaba como una curiosidad técnica o una herramienta para la optimización de procesos rutinarios, hoy se entiende como una fuerza disruptiva que afecta a la esencia misma de la autoridad periodística. En este sentido, la literatura actual ha dejado de lado los relatos deterministas que vaticinaban un reemplazo masivo de periodistas por algoritmos, para centrarse en modelos de colaboración que Lindén (2020) o Tejedor y Vilà (2021) conceptualizan como *augmented journalism* o exo-periodismo. Este giro hacia la complementariedad sugiere que la IA actúa como una prótesis cognitiva que expande las capacidades del profesional, permitiéndole gestionar volúmenes de datos inabarcables para la mente humana sin por ello diluir su agencia ética (Wu et al., 2019; Bakke y Barland, 2022). Como sostienen Nurlatifah e Irwansyah (2019), el algoritmo es capaz de detectar patrones e inconsistencias a gran velocidad, pero carece de la sensibilidad contextual y del juicio interpretativo necesarios para la toma de decisiones editoriales complejas, facultades que permanecen —y deben permanecer— en el dominio de lo humano (Deuze & Beckett, 2022; Kothari & Hickerson, 2020).

Esta evolución histórica no ha sido un proceso lineal. Si bien organizaciones como *The Big Ten Network* o *The Los Angeles Times* ya utilizaban sistemas automatizados para reportes deportivos o actualizaciones sísmicas hace más de una década (Danzon-Chambaud, 2021; Sánchez-García et al., 2023), la democratización de la IA generativa a finales de 2022 ha marcado un punto de inflexión radical. El ascenso vertiginoso de herramientas como ChatGPT, que alcanzó niveles de adopción masiva en tiempos récord (Milmo, 2023), ha empujado al periodismo a una "Cuarta Revolución Industrial" (Micó et al., 2022; Dhiman, 2023). No obstante, este avance técnico convive con una preocupación creciente por la integridad del mensaje y la validez del conocimiento producido. Investigaciones recientes señalan que los Modelos de Lenguaje de Gran Escala (LLMs), contruidos sobre arquitecturas tipo *transformer*, presentan una dificultad intrínseca para ser auditados internamente, lo que limita la capacidad de los medios para verificar los fundamentos de las respuestas generadas (Patil y Gudivada, 2024; Liang et al., 2024). Autores como Shao et al.

(2024) o Gao et al. (2023) advierten que la fluidez sintáctica de estos modelos suele enmascarar una fragilidad semántica que conduce a la simulación de fuentes y a la fabricación de datos, un fenómeno que compromete la veracidad que sustenta la confianza social en el periodismo y altera los marcos tradicionales de autoridad epistémica (Baryshnikov, 2024; Lindemann, 2023).

En este contexto, la paradoja de la desinformación se sitúa en el centro del escenario académico. La literatura documenta una dualidad crítica: mientras la IA fortalece la lucha contra los bulos mediante algoritmos de *fact-checking*, *random forests* y sistemas de monitorización avanzada (Arias Jiménez et al., 2022) también se convierte en el motor más eficiente para la creación de contenidos manipulados a gran escala. La capacidad de los LLMs para clonar voces, producir imágenes hiperrealistas y generar relatos ficticios con una verosimilitud asombrosa ha erosionado la credibilidad de las instituciones informativas (Morato y Nunes, 2023; Muñoz-Vela, 2024). Estudios de Newsguard (2023) alertan sobre la proliferación de cientos de portales de noticias sintéticas que operan sin supervisión humana, desestabilizando el espacio público con narrativas desinformativas que tienen un 70% más de probabilidades de volverse virales que la información verídica (Vosoughi et al., 2018). Esta "paradoja de la transparencia" (Blázquez-Ruiz, 2022) revela que el periodismo habita hoy un entorno donde las mismas herramientas que garantizan la verificación pueden ser utilizadas para erosionar la realidad, lo que obliga a la profesión a transitar del rol de *gatekeeper* tradicional al de un curador crítico de contenidos o *gatewatcher* (Bruns, 2003; Loo-Vázquez et al., 2016).

Esta tensión se agrava ante la opacidad intrínseca de los sistemas actuales. La auditoría algorítmica ha emergido como una herramienta clave, con modelos estructurados como PRISM o la estrategia *human-in-the-loop* (Mokander et al., 2023; Zheng et al., 2025). Sin embargo, estas auditorías revelan un patrón persistente de escasa trazabilidad y propensión a la fabricación de fuentes (Chiang y Lee, 2024; Wu et al., 2024). La dificultad de comprender cómo una entidad aparentemente servicial resulta internamente inescrutable se define como una "inescrutabilidad inmanente" (Masís-González, 2024; Cuéllar-Rodríguez, 2023). Ello ha motivado la urgencia de marcos regulatorios como la Ley de Inteligencia Artificial de la Unión Europea (AI Act), analizada por Novelli et al. (2024), la cual establece criterios obligatorios de responsabilidad, ciberseguridad y transparencia, clasificando los sistemas según su nivel de riesgo para proteger derechos fundamentales en sectores críticos.

Un eje fundamental que esta tesis explora, y que la literatura internacional ha comenzado a abordar bajo el prisma de la justicia de datos, es el desafío de las lenguas minorizadas y los contextos periféricos. Investigaciones de Ermolova et al. (2024) y Ta y Lee (2023) denuncian que los modelos entrenados mayoritariamente en lenguas dominantes generan asimetrías lingüísticas que perjudican sistemáticamente a las comunidades no hegemónicas. Como señalan Agarwal et al. (2024), el idioma en el que se formula una consulta no es un detalle trivial; afecta

directamente al razonamiento ético y a la calidad de la respuesta del modelo. Esta desigualdad lingüística no solo es un déficit técnico, sino una forma de invisibilización de narrativas locales y regionales. En entornos como el español, donde el 77,5% de los municipios se enfrentan al riesgo de convertirse en desiertos informativos (Negreira-Rey et al., 2023), la falta de presencia digital de estas comunidades provoca su exclusión de los corpus de entrenamiento de los LLMs. Se produce así una retroalimentación de la invisibilidad: lo que no está en los datos masivos, no existe para la IA, y lo que no existe para la IA, tiende a desaparecer de la esfera pública globalizada (Mihajlov-Prokopović et al., 2019).

Este riesgo de homogeneización cultural refuerza la necesidad de una soberanía digital que Kasperian (2021) y Peña-Fernández et al. (2023) sitúan como una condición indispensable para la justicia algorítmica. El euskera, en este marco, emerge como un caso de estudio paradigmático. Pese a los esfuerzos institucionales, la lengua vasca continúa siendo marginal en los desarrollos comerciales de gran escala, lo que genera una vulnerabilidad epistémica que afecta tanto a la recuperación de información como a la generación de contenidos contextualizados. La literatura científica insiste en que la inclusión lingüística no debe ser una concesión comercial, sino un derecho democrático que garantice la pluralidad informativa. De ahí que la decisión de muchas redacciones de adoptar una postura de "cautela informativa" se interprete como una respuesta ética necesaria ante herramientas que, en su configuración actual, no garantizan la precisión ni el respeto por las especificidades territoriales, un patrón observado en la reticencia de los profesionales a delegar la selección y edición de contenidos (Wu et al., 2018).

Por otro lado, la incorporación de la IA en el periodismo está transformando de manera irreversible las estructuras empresariales y los perfiles profesionales. La emergencia de figuras híbridas como el periodista de datos, el curador algorítmico o el *journalist-programmer* demanda un cambio de paradigma en la formación (Lopezosa et al., 2023; Túniz-López et al., 2021; Pinto & Barbosa, 2024). No obstante, existe un desfase preocupante entre la rapidez de la innovación técnica y la lentitud de las estructuras educativas. Diversos autores señalan que la integración de competencias en IA en los planes de estudio sigue siendo incipiente y fragmentada (Connor y Cali, 2024). Esta brecha formativa deja a los profesionales en una situación de vulnerabilidad, obligándoles a recurrir al autoaprendizaje y a depender acríticamente de tecnologías cuya lógica interna desconocen (Sonni et al., 2024; Jones et al., 2022). Sin una alfabetización algorítmica profunda —que abarque dimensiones técnicas, críticas y éticas (Ng et al., 2021)— el periodismo corre el riesgo de perder su función de vigilancia social, convirtiéndose en un mero distribuidor de contenidos generados por infraestructuras externas sobre las que posee escaso control editorial.

Este desplazamiento de la soberanía hacia las grandes plataformas tecnológicas (Google, Meta, OpenAI) es uno de los dilemas más profundos que afronta la profesión (Simon, 2022; Broussard et al., 2019). La mediación algorítmica determina qué

contenidos se amplifican y cuáles se invisibilizan, a menudo priorizando lógicas comerciales sobre el valor informativo o el interés público (Luitse y Denkena, 2021). Ante esta situación, la literatura reclama una gobernanza proactiva que trascienda la mera regulación técnica. Si bien el uso de IA puede aliviar la precaria situación económica de muchas empresas mediáticas (Kim y Kim, 2017), los altos costes de desarrollo de aplicaciones propias suelen derivar en una dependencia aún mayor de los gigantes tecnológicos. Los investigadores coinciden en que la verdadera salvaguarda reside en la supervisión humana constante y en el compromiso de los medios con principios de transparencia y rendición de cuentas, informando al público sobre cuándo y cómo se emplea la IA en la creación de contenidos (Ufarte-Ruiz, Calvo Rubio & Murcia-Verdú, 2021; Sanahuja & Esteban, 2023).

Adicionalmente, el impacto de la IA en la polarización social y política es un eje de preocupación creciente. El uso de *microtargeting* y la exposición selectiva en redes sociales han contribuido a la creación de burbujas informativas y cámaras de eco que limitan el acceso de los usuarios a perspectivas diversas (Pariser, 2011; Cinelli et al., 2021). El despliegue de *socialbots* diseñados para difundir desinformación masiva añade una capa de complejidad, consolidando narrativas falsas que moldean la opinión pública (Boshmaf et al., 2011; Rossetti & Zaman, 2023). Este fenómeno se ha vuelto especialmente visible en contextos electorales recientes, donde la desinformación ha erosionado la confianza en las instituciones democráticas y mediáticas (Casero-Ripollés et al., 2023; Edelman, 2023). La confianza mediática, por tanto, depende no solo de la calidad del producto final, sino de la transparencia de los procedimientos utilizados para garantizar su precisión (Kohring & Matthes, 2007; Strömbäck et al., 2020).

Finalmente, el estado del conocimiento revela que la IA es, en última instancia, un artefacto atravesado por estructuras de poder y decisiones ideológicas (Castillo-Eslava et al., 2023; Luitse y Denkena, 2021). No existen sistemas neutrales; cada algoritmo hereda los sesgos de sus creadores y de los datos con los que ha sido alimentado, reproduciendo estereotipos demográficos y de género (Liu et al., 2024; Algaba et al., 2024). En un ecosistema como el vasco, donde la ética y la identidad profesional están profundamente enraizadas en el territorio, la transición tecnológica debe ser gobernada desde una perspectiva situada y crítica. El papel de la alfabetización mediática e informacional (AMI) resulta aquí fundamental, no solo para fortalecer la capacidad crítica individual, sino para complementar las herramientas tecnológicas de verificación (Amazeen & Bucy, 2019; Vraga et al., 2022). Solo a través de un enfoque multidisciplinario que combine la formación ética, la sensibilidad lingüística y la regulación de las prácticas de IA será posible aprovechar su potencial innovador sin comprometer los pilares fundamentales del periodismo. Esta tesis se inserta precisamente en ese vacío, aportando una mirada empíricamente fundamentada que visibiliza los desafíos específicos de un ecosistema que busca su lugar en la era algorítmica sin renunciar a su propia diversidad cultural y lingüística.

De este modo, se concluye que el éxito de la IA en el periodismo no se medirá por la sofisticación de sus algoritmos, sino por la capacidad de los profesionales para integrarlos bajo marcos de responsabilidad social. La convergencia entre IA y comunicación abre oportunidades inéditas para repensar el futuro de la profesión, siempre y cuando se mantengan los principios éticos que han definido históricamente esta labor en la sociedad contemporánea.

5. Metodología

5.1. Diseño de investigación

La arquitectura metodológica de esta tesis doctoral se fundamenta en un paradigma de investigación mixto, diseñado para captar la complejidad de la IA en su intersección con el periodismo. Lejos de entender la tecnología como un fenómeno puramente técnico, esta investigación la aborda como un proceso sociotécnico que transforma las rutinas, la ética y las estructuras de poder en los medios de comunicación. Por ello, se ha optado por una triangulación metodológica que permite contrastar la dimensión estadística (encuesta), la densidad interpretativa (entrevistas) y el rendimiento empírico de los sistemas (experimentación con LLMs).

El diseño se estructura como una investigación multimétodo aplicada a un estudio de caso múltiple: el ecosistema mediático del País Vasco. La elección de este marco geográfico y lingüístico responde a su singularidad como territorio con lengua propia y una estructura mediática diversificada (pública, privada y comunitaria). La investigación se articula sobre cuatro pilares que garantizan la validez interna y externa de los resultados:

Tabla 11: Marco estratégico de la investigación

Dimensión	Técnica	Fundamentación	Aplicación
Cuantitativa	Encuesta transversal	Creswell (2014): Representatividad y generalización de tendencias.	Percepción de 501 periodistas vascos sobre el impacto de la IA.
Cualitativa	Entrevistas semiestructuradas	Kvale (1996): Acceso a la lógica interna de las organizaciones.	6 responsables estratégicos de medios líderes en Euskadi.
Experimental	Auditoría de LLMs	Bender et al. (2021): Evaluación crítica de sesgos y opacidad técnica.	Interacción controlada con ChatGPT, Gemini, Claude, Copilot y Perplexity.
Documental	Revisión sistemática	Hine (2015): Etnografía digital y marcos regulatorios.	Análisis de guías éticas y literatura científica (2023-2025).

Fuente: Elaboración propia

El componente cuantitativo tiene como objetivo cuantificar el conocimiento, el uso y las actitudes de los profesionales ante el fenómeno de la automatización. El universo de estudio está integrado por periodistas y comunicadores en activo en la Comunidad Autónoma Vasca.

- **Muestreo y representatividad:** Se obtuvo una muestra de 501 respuestas válidas (n=501). Tomando como referencia el censo estimado de 5.000 profesionales en el sector (Gobierno Vasco, 2022), el margen de error para un nivel de confianza del 95% es de $\pm 4,15\%$.
- **Procedimiento:** La encuesta se administró de forma online (CAWI) con apoyo telefónico entre mayo y junio de 2024. Se utilizó la *Guía de Comunicación Abierta del Gobierno Vasco* y la base de datos de la *Asociación Vasca de Periodistas* para asegurar que los participantes fueran profesionales en activo.
- **Segregación de datos:** La muestra presenta un equilibrio de género (55,1% hombres; 44,5% mujeres; 0,4% otros). Se analizaron variables de control como la edad, los años de experiencia, la tipología del medio y, fundamentalmente, el nivel de responsabilidad (redactores vs. directivos), lo que permite identificar si la adopción de la IA es una imposición *top-down* o una demanda *bottom-up*.

Para comprender la gobernanza algorítmica y las políticas de innovación, se realizaron seis entrevistas semiestructuradas con responsables estratégicos. Estas entrevistas, de una duración de entre 30 y 50 minutos, se llevaron a cabo en abril y mayo de 2024 mediante videoconferencia y encuentros presenciales. Los perfiles seleccionados representan el espectro completo del ecosistema mediático vasco:

Tabla 10: Registro de entrevistas en profundidad (Informantes clave)

Cód.	Informante	Organización	Perfil / Cargo
I1	Iván Pérez	Vocento / El Correo	Responsable de nuevas tendencias tecnológicas y calidad.
I2	Jon Telletxea	EiTB Media	Subdirector de la radiotelevisión pública vasca.
I3	Javier Turrado	Grupo Noticias / Deia	Director del área tecnológica.

I4	Beñat Zaldúa	Naiz	Responsable de contenidos y planificación estratégica.
I5	Pello Urcelai	Berria	Responsable de TI en el principal diario en euskera.
I6	Jon Egaña	Cadena SER / Radio Bilbao	Jefe de Redacción (perspectiva de radio y proximidad).

Fuente: Elaboración propia

Cada una de las publicaciones que integran esta tesis responde a un diseño específico que contribuye a la visión global del fenómeno:

- **Artículo 1 (Transdigital):** Adopta un diseño exploratorio-descriptivo. Combina la encuesta (n=501) con las entrevistas cualitativas para captar tanto la dimensión estadística del uso de la IA como las interpretaciones institucionales que orientan su adopción. Es el artículo que sienta las bases de la transparencia y la ética en las redacciones vascas.
- **Artículo 2 (Hipertext.net):** Se estructura como un experimento comparativo de auditoría algorítmica. Se interactuó con cinco modelos (ChatGPT, Gemini, Claude, Copilot y Perplexity) utilizando un corpus de preguntas sensibles sobre la actualidad de Euskadi y el País Vasco. Se incluyeron "consultas de ciudadana" (perspectiva de Bilbao) para evaluar la recuperación de información local. Las variables analizadas incluyeron la trazabilidad de fuentes, la extensión y la diversidad epistémica.
- **Artículo 3 (Frontiers in Communication):** Emplea un diseño centrado en la brecha formativa. Utiliza los datos de la encuesta y entrevistas para analizar las tensiones éticas (autoría, veracidad y confianza social) derivadas de la falta de capacitación específica en IA generativa entre los profesionales de diferentes rangos de edad.
- **Artículo 4 (Estudios del Mensaje Periodístico):** Se define como un estudio de caso múltiple comparado. Analiza las estrategias de EitB, El Correo, DEIA, Berria y Cadena SER. Mediante la triangulación de entrevistas y revisión documental, identifica la "cautela informativa" como un modelo de gobernanza frente a la adopción acelerada de tecnologías externas.

- **Artículo 5 (Mediatika):** Consiste en una revisión sistemática y crítica de la literatura. Analiza el papel que tienen los *deepfakes* y los LLMs en la producción de desinformación, evaluando las herramientas de verificación que los periodistas tienen a su disposición en el contexto actual de la desinformación amplificada por IA.
- **Artículo 6 (Interconectando Saberes):** Propone un diseño teórico-crítico. A través del análisis de casos documentados, el artículo examina cómo el entrenamiento de los modelos de IA puede reproducir sesgos de género y desigualdades estructurales, aportando un enfoque interpretativo indispensable para la ética periodística.

Tabla 11: Matriz de técnicas aplicadas por artículo

Técnica	P1	P2	P3	P4	P5	P6
Encuesta (n=501)	X		X	X		
Entrevistas (I1-I6)	X		X	X		
Experimento LLMs		X				
Análisis de contenido		X		X		X
Revisión documental					X	X

Fuente: Elaboración propia

Finalmente, la metodología se caracteriza por su pluralidad epistémica y lingüística. El uso de materiales y fuentes en castellano, euskera e inglés responde a la realidad multilingüe del campo de estudio. El proceso de investigación ha seguido estrictos protocolos éticos: anonimización de las respuestas de la encuesta, consentimiento

informado en las entrevistas y una transparencia total en los *prompts* utilizados en la fase experimental, asegurando que los resultados sean trazables y replicables. Esta robustez metodológica garantiza que las conclusiones de la tesis no sean meras observaciones aisladas, sino el resultado de un contraste riguroso entre la práctica profesional y el rendimiento tecnológico.

6. RESULTADOS

Esta sección articula los hallazgos de la investigación en respuesta a los objetivos específicos planteados. El análisis se presenta como una síntesis integradora de las evidencias obtenidas en el compendio de publicaciones, triangulando la encuesta a 501 profesionales, las entrevistas a responsables estratégicos y la experimentación técnica. Los resultados confirman que el ecosistema mediático vasco atraviesa una fase de transición hacia la automatización definida por la asimetría técnica y la prioridad absoluta del control editorial sobre la velocidad de producción.

Tabla 124: Matriz de triangulación de hallazgos, evidencias y producción científica

Dimensión / Eje temático	Hallazgo principal	Evidencia empírica clave	Artículos de referencia
Rutinas periodísticas	Integración asistencial bajo el modelo de "cautela informativa"; la IA apoya, no sustituye.	Estudios de caso en EiTB (Podcast <i>Ahalaken</i>) y SER Euskadi.	A1, A4
Formación y alfabetización	Brecha formativa crítica que genera inseguridad operativa y dependencia del autoaprendizaje.	Encuesta a 501 periodistas: solo el 14,1% posee formación reglada.	A1, A3
Lengua y contexto regional	Bajo rendimiento de los LLMs en euskera y riesgo de invisibilización del contexto local.	Experimentos con 5 modelos de lenguaje (ChatGPT, Gemini, etc.) sobre noticias vascas.	A2
Ética y gobernanza	Alta preocupación profesional frente a la ausencia de protocolos institucionales y guías de estilo.	Encuesta (63,91% de freno ético) y entrevistas a responsables de innovación.	A1, A4
Desinformación	Emergencia de herramientas de verificación para contrarrestar contenidos sintéticos y <i>deepfakes</i> .	Análisis de herramientas de <i>fact-checking</i> y revisión sistemática de literatura.	A5
Sesgos de género	Reproducción de estereotipos y reducción de la visibilidad femenina en los <i>outputs</i> algorítmicos.	Análisis algorítmico y crítico de la justicia de datos y representación.	A6

Fuente: Elaboración propia

6.1. Efectos de la implementación de la IA en las dinámicas laborales y editoriales (OE1)

La incorporación de la IA generativa en el País Vasco no está produciendo una sustitución del periodista, sino una especialización de sus funciones hacia tareas de supervisión y validación. La tecnología se emplea bajo un modelo de "cautela informativa", actuando como un sistema de asistencia para procesos técnicos y administrativos de bajo riesgo, mientras que la decisión editorial permanece bajo un estricto criterio humano.

La solidez de este hallazgo se sustenta a través de los siguientes indicadores extraídos de la investigación:

- **Frecuencia de uso y brecha por soportes:** El 45,91% de los periodistas vascos aún no ha integrado herramientas de IA en su rutina diaria, frente a un escaso 7,19% que las emplea con regularidad cada jornada. Esta adopción es profundamente desigual según el soporte: los medios digitales lideran la integración con un 50% de uso semanal y solo un 10,5% de rechazo total, cifras que contrastan con la prensa escrita, donde el 54,20% de la plantilla nunca utiliza esta tecnología, tal como se extrae de los resultados del artículo aceptado en *Transdigital*.
- **Distribución técnica de tareas:** El uso de la IA es marcadamente accesorio y complementario. Las funciones más habituales son la traducción automática (46,4%), las consultas temáticas de datos (30%), la corrección ortotipográfica (27,6%) y la transcripción de audio (23%). En cambio, el uso de la IA para la generación directa de contenido informativo es muy limitado, con un 21% global que cae hasta el 13% en soportes tradicionales, según los datos detallados en la revista *Estudios sobre el Mensaje Periodístico*.
- **Diferenciación por roles y departamentos:** Existe una clara asimetría en la apropiación tecnológica según el puesto de trabajo. Los profesionales de áreas técnicas (60,5%) y comunicación corporativa (64,7%) utilizan la IA casi el doble que los redactores (37,4%) o locutores (26,3%), lo que confirma que la tecnología penetra en la periferia de la redacción antes que en el núcleo creativo, hallazgo que se desprende de las investigaciones en *Frontiers in Communication* y *Estudios sobre el Mensaje Periodístico*.
- **Hitos de experimentación y formatos emergentes:** La innovación en el territorio se manifiesta en proyectos singulares de expansión narrativa, como la clonación de la voz del meteorólogo Peio Zabala para el podcast *Ahalaken* o la síntesis vocal de Kepa Junkera para Radio Bilbao. Estos casos demuestran que la IA se utiliza para la recuperación de la memoria cultural y la creación multimedia bajo protocolos de consentimiento y supervisión multidisciplinar, como se analiza en los artículos de *Transdigital* y *Estudios sobre el Mensaje Periodístico*.

- **Vigilancia ante el sesgo y la desinformación:** La experimentación técnica revela que los modelos globales presentan una infrarrepresentación sistemática de fuentes regionales; de las 129 fuentes analizadas en casos locales, solo el 15,5% eran medios vascos, lo que obliga al periodista a actuar como un "curador algorítmico" para corregir la invisibilización territorial y los sesgos de género detectados, tal como se extrae de los trabajos en *Hipertext.net*, *Interconectando Saberes* y *Mediatika*.

Desde una perspectiva operativa, la 'cautela informativa' identificada en esta investigación se traduce en una metodología de trabajo diaria. Los medios vascos han integrado la IA no como un fin, sino como una herramienta de validación técnica paso a paso. El periodista aplica esta cautela como un protocolo de control de calidad en su rutina para asegurar que los procesos de traducción o transcripción no comprometan el rigor de la pieza final antes de su publicación

6.2. Percepciones, actitudes y niveles de alfabetización tecnológica (OE2)

La investigación revela una paradoja estructural en el ecosistema mediático vasco: existe un elevado interés teórico y pragmático por las potencialidades de la IA, pero este convive con un déficit formativo sistémico que genera inseguridad operativa. Esta brecha de capacitación, que afecta especialmente a los perfiles más vinculados a la redacción, actúa como el principal freno para una integración ética, derivando en una dependencia del autoaprendizaje y en una percepción de la tecnología como una "caja negra" difícil de auditar profesionalmente.

La profundidad de este diagnóstico se apoya en los siguientes indicadores cuantitativos y cualitativos:

- **Nivel de capacitación reglada:** Solo el 14,1% de los 501 profesionales encuestados afirma haber recibido algún tipo de formación específica en IA. Esta carencia es más acusada en los roles puramente periodísticos, como redactores (13,8%) y presentadores (6,3%), frente a perfiles técnicos (66,7%) o de gestión (25%), tal como se detalla en los resultados de *Frontiers in Communication*.
- **Brecha por tipo de medio y alcance:** Los profesionales de medios nativos digitales presentan tasas de formación superiores (25,8%) a los de soportes tradicionales como la prensa (9,6%), la radio (10,8%) o la televisión (5,8%). Asimismo, los trabajadores de medios con difusión internacional reportan casi el doble de formación (21,7%) que los de medios locales (12,2%), lo que sugiere una polarización tecnológica vinculada al músculo financiero de la organización, según se extrae de *Frontiers in Communication*.

- **Origen de la formación y barreras de acceso:** Únicamente el 8,5% de los periodistas ha recibido formación por parte de su empresa, lo que obliga a que un 12,9% busque capacitación de manera independiente y la gran mayoría dependa del autoaprendizaje. Las barreras citadas para no formarse incluyen la falta de recursos en medios locales, la excesiva carga laboral y el desconocimiento de fuentes de calidad, hallazgos que se desprenden de *Frontiers in Communication y Transdigital*.
- **Actitudes y motivación profesional:** Una amplia mayoría de los periodistas (71,26%) identifica la IA como una oportunidad para automatizar tareas repetitivas y liberar tiempo para el trabajo creativo. Sin embargo, esta visión optimista convive con un 55,89% de profesionales que teme un incremento significativo de la desinformación y un 35,73% que alerta sobre la exacerbación de sesgos digitales, hallazgos que se documentan detalladamente en el artículo aceptado en *Transdigital*.
- **Barreras éticas y generacionales:** Las preocupaciones éticas son el principal motivo de reticencia para el 63,91% de quienes no usan IA. Este recelo es mayor en profesionales con más de 20 años de experiencia (47,62%) frente a los noveles con menos de 5 años (15,65%), lo que evidencia una brecha generacional en la percepción del riesgo para la credibilidad y la autoría, según los datos de *Transdigital*.
- **Necesidad de alfabetización crítica y de género:** La alfabetización actual se percibe como meramente instrumental, ignorando dimensiones esenciales. Existe una demanda urgente de formación que incluya la detección de sesgos de género y la transparencia algorítmica, ya que la falta de esta mirada crítica permite que la IA perpetúe estereotipos patriarcales en la esfera pública, tal como se analiza en *Interconectando Saberes y Mediatika*.

La investigación concluye que el "abismo formativo" detectado no es solo un déficit de habilidades técnicas, sino una amenaza a la autonomía profesional. Al depender mayoritariamente del autoaprendizaje y la experimentación informal, el periodista vasco carece de las herramientas necesarias para ejercer una supervisión consciente (*human-in-the-loop*). Esta situación genera una posición de vulnerabilidad frente a las plataformas globales: el profesional sabe que debe innovar, pero no posee la capacidad técnica para auditar los resultados del algoritmo.

Por tanto, la formación no puede seguir siendo entendida como un barniz operativo, sino como un requisito para el ejercicio ético del periodismo y la comunicación en general. La transición de un uso "asistencial" a uno "estratégico" de la IA en Euskadi depende de institucionalizar programas de capacitación que vinculen la destreza técnica con la responsabilidad social. En última instancia, la soberanía del periodismo vasco frente a la automatización solo será posible si el profesional recupera el control sobre la "caja negra" mediante una alfabetización profunda que garantice la trazabilidad y la equidad en el discurso público.

6.3. Eficacia, adecuación contextual y barreras en entornos multilingües (OE3)

La investigación identifica una brecha de "centralismo algorítmico" en el desarrollo de la IA generativa, donde las herramientas globales muestran una degradación severa de su eficacia al operar en contextos regionales y en lengua vasca. Se observa que la precisión informativa y la trazabilidad de los modelos son directamente proporcionales a la escala geográfica del suceso: mientras que el rendimiento es óptimo en temas de alcance nacional, se producen "alucinaciones" y omisiones sistémicas al abordar la realidad informativa de proximidad en Euskadi.

La profundidad de este diagnóstico técnico se sustenta en los siguientes indicadores extraídos de la experimentación y el análisis documental:

- **Sesgo territorial en la producción textual:** Los modelos generativos muestran una capacidad desigual para gestionar la información local. Perplexity y ChatGPT lideran la extensión narrativa con 680 y 638 palabras de promedio respectivamente en noticias de actualidad, mientras que modelos como Gemini ofrecen respuestas residuales de apenas 96 palabras, incurriendo en fórmulas evasivas ante consultas sobre Euskadi. De las 129 fuentes citadas en el análisis de casos regionales, el 67,4% fueron de ámbito nacional, frente a un escaso 15,5% de fuentes locales o regionales, lo que evidencia una infrarrepresentación de medios como Eitb, Berria o Deia, tal como se extrae de *Hipertext.net*.
- **Limitaciones de trazabilidad y verificación:** La transparencia documental decae drásticamente en la escala regional. Solo ChatGPT y Perplexity mantienen una trazabilidad completa (100% de enlaces activos), mientras que Claude, Copilot y Gemini presentan una opacidad casi total en las referencias locales, utilizando fórmulas vagas como "según diversos medios". De las 111 fuentes mencionadas en consultas de proximidad, solo el 54% resultaron verificables, según los datos de *Hipertext.net*.
- **La barrera del euskera y el reconocimiento de voz:** La integración de la IA en tareas de edición encuentra obstáculos morfosintácticos y dialectales. En herramientas de transcripción integradas, como la plataforma *Guau* de Eitb, se detectan dificultades de calidad en función del dialecto de euskera utilizado, lo que obliga a una revisión manual constante. Asimismo, medios como *Berria* han tenido que suspender temporalmente ensayos de transcripción automática por falta de precisión lingüística, según se desprende de *Estudios sobre el Mensaje Periodístico*.
- **Desarrollo de infraestructuras propias:** Ante la ineficacia de los modelos globales, el ecosistema vasco está impulsando soluciones situadas. Destaca el consorcio liderado por *Berria* junto a otros medios en euskera (Hamaika

Telebista o Euskalerrria Irratia) para desarrollar traductores neuronales y sintetizadores de voz específicos que respeten las particularidades del idioma, hallazgos que se recogen en *Estudios sobre el Mensaje Periodístico*.

- **Vulnerabilidad ante la desinformación multilingüe:** La falta de *datasets* robustos en euskera limita la eficacia de las herramientas de *fact-checking* automatizado. La IA generativa facilita de este modo la creación de *deepfakes* y textos sintéticos con una verosimilitud superior a la capacidad de detección de los sistemas actuales en entornos de lenguas minorizadas, lo que incrementa el riesgo de manipulación informativa en el territorio, tal como se analiza en *Mediatika*.

Los resultados confirman que la IA, en su configuración actual, actúa como un agente de "homogeneización cultural". Al priorizar fuentes centralizadas y marcos de referencia nacionales, los modelos globales invisibilizan la agenda local vasca. Esta investigación concluye que la "cautela informativa" mencionada anteriormente es una respuesta técnica necesaria: si los periodistas delegaran la producción en estas herramientas, se produciría una "erosión del pluralismo territorial".

Por tanto, el éxito de la IA en Euskadi no depende de la adopción de modelos externos, sino del reclamo de una "agencia algorítmica propia". La soberanía tecnológica en el País Vasco exige el desarrollo de modelos entrenados con corpus lingüísticos locales para garantizar que la automatización no degrade la calidad informativa ni la representación de la diversidad lingüística en la esfera digital.

6.4. Identificación de sesgos de género y desigualdad algorítmica (OE4)

La investigación concluye que la IA informativa en el contexto actual no actúa como un agente neutral, sino como un sistema que cristaliza y amplifica los sesgos de género preexistentes. Los algoritmos, al ser entrenados con *datasets* que arrastran desigualdades históricas, tienden a reproducir estereotipos que invisibilizan a las mujeres en roles de autoridad y las relegan a representaciones tradicionales, lo que supone un riesgo directo para el pluralismo y la equidad en la esfera pública vasca. La profundidad de este análisis se sustenta en la evidencia documental y empírica extraída de las diversas fases del estudio:

- **Percepción del riesgo de sesgo:** Un 35,73% de los periodistas vascos considera que la IA incrementará significativamente los sesgos digitales y las desigualdades sociales, mientras que un 36,73% adicional prevé un impacto moderado. Esta preocupación es transversal, aunque los hombres manifiestan una inquietud ligeramente superior (52,38%) frente a las mujeres (46,94%) respecto a las barreras éticas generales, según los resultados obtenidos en el artículo de *Transdigital*.

- **Segregación temática y de fuentes:** Los sistemas algorítmicos tienden a priorizar fuentes masculinas para la cobertura de "hard news" (política, economía o tecnología), mientras que las representaciones femeninas se concentran en ámbitos vinculados al entretenimiento o los cuidados. Esta "segregación algorítmica" es el resultado de una falta de diversidad en los datos de entrenamiento que los modelos de lenguaje (LLMs) replican de forma opaca, tal como se extrae de *Interconectando Saberes*.
- **Feminización de la asistencia técnica:** Se observa un patrón de género en el diseño de interfaces; la mayoría de los asistentes virtuales (Siri, Alexa, Cortana) emplean voces femeninas por defecto, reforzando la asociación de la mujer con roles de servidumbre y asistencia dócil. Por el contrario, sistemas orientados a contextos profesionales o de autoridad, como Watson de IBM, suelen utilizar voces masculinas, hallazgo que se detalla en el artículo de la revista *Interconectando Saberes*.
- **Desinformación de género y acoso:** La IA facilita nuevas formas de violencia digital, como la creación de *deepfakes* pornográficos sin consentimiento ("porno de venganza") o el uso de *socialbots* para campañas de desprestigio contra lideresas políticas y periodistas. Estas prácticas buscan deslegitimar la autoridad de la mujer y limitar su participación en el debate público, según se analiza en *Interconectando Saberes* y *Mediatika*.
- **Impacto en la visibilidad laboral:** Los algoritmos de recomendación y búsqueda a menudo presentan sesgos en la jerarquización de perfiles; por ejemplo, las búsquedas de términos como "CEO" devuelven mayoritariamente imágenes de hombres, mientras que los anuncios de empleo automatizados tienden a mostrar salarios más altos a usuarios identificados como masculinos, evidencia documentada en *Interconectando Saberes*.
- **Necesidad de una alfabetización de género:** Ante esta realidad, los profesionales señalan que la formación técnica es insuficiente si no se acompaña de una capacitación en perspectiva de género. Resulta crítico que solo el 14,1% de los periodistas posea formación en IA, ya que esta carencia impide detectar y corregir los sesgos antes de que el contenido llegue a la audiencia, según se desprende de *Frontiers in Communication* y *Interconectando Saberes*.

La investigación demuestra que la automatización sin auditoría ética actúa como un "espejo deformante" de la sociedad. En Euskadi, un ecosistema que busca activamente la paridad, la delegación de la producción informativa en modelos globales no auditados supone un retroceso en términos de representación. La IA no solo hereda el machismo de los datos históricos, sino que lo "normaliza" bajo una apariencia de objetividad técnica.

Se concluye, por tanto, que la lucha por la igualdad en el periodismo del siglo XXI no se libra únicamente en las estructuras de las redacciones, sino en la justicia de datos.

Resulta imperativo transitar hacia una "gobernanza de género" en el desarrollo de la IA, donde la supervisión humana actúe como un filtro crítico que impida la consolidación de estos estereotipos en el código algorítmico. Sin esta mirada, la innovación tecnológica corre el riesgo de convertir la esfera digital en un espacio de exclusión para las mujeres. Esta problemática conecta directamente con el déficit de capacitación analizado en el OE2; la investigación demuestra que la invisibilización de la mujer en los resultados de la IA no podrá corregirse mientras persista el abismo formativo actual (donde solo el 14,1% posee formación reglada), ya que el profesional carece de las competencias críticas necesarias para auditar y neutralizar estos sesgos en la fase de producción

6.5. Transparencia, gobernanza y rendición de cuentas (OE5)

La investigación revela una asincronía estructural entre la velocidad de adopción tecnológica y la formalización de marcos regulatorios internos en los medios vascos. Se observa una ausencia casi total de mecanismos institucionalizados de transparencia y rendición de cuentas (*accountability*), lo que sitúa la responsabilidad ética en una "zona de indeterminación normativa". En este escenario, la gestión de los riesgos algorítmicos no depende de directrices corporativas explícitas, sino del criterio individual y la autorregulación voluntarista de los profesionales.

La solidez de este diagnóstico se apoya en los siguientes indicadores transversales:

- **Inseguridad normativa y barreras éticas:** El 63,91% de los periodistas identifica las dimensiones éticas como el principal factor limitante para integrar la IA en su desempeño. Esta percepción de riesgo se vincula directamente con la falta de normativas claras y el temor a que una comunicación deficiente sobre el uso de algoritmos genere rechazo en el público, tal como se extrae de los resultados publicados en *Transdigital*.
- **Ausencia de protocolos y guías de estilo:** A pesar del interés por la tecnología, la respuesta institucional es mínima. La práctica inexistencia de Libros de Estilo actualizados o de guías de gobernanza algorítmica específicas convierte la rendición de cuentas en una práctica informal. Los responsables de medios reconocen que, aunque existen decálogos institucionales incipientes (como el de EITB), la mayoría de las redacciones carece de un sistema de auditoría que garantice la trazabilidad, según se desprende de *Estudios sobre el Mensaje Periodístico*.
- **Demanda de transparencia hacia la audiencia:** Existe un consenso profesional sobre la necesidad de informar al público: el 55,09% de los periodistas considera imprescindible comunicar explícitamente el uso de IA en cualquier etapa del proceso informativo, mientras que un 33,13% cree que debe hacerse al menos cuando la IA actúa como fuente principal, de acuerdo con los datos de *Transdigital*.

- **Riesgo de erosión de la confianza mediática:** La falta de claridad sobre el uso de sistemas automatizados es percibida como un vector de desconfianza. El 55,89% de los profesionales cree que la IA incrementará los riesgos de desinformación, lo que obliga a las empresas a transitar de una "cautela reactiva" a una "transparencia proactiva" para mantener el contrato con la audiencia, hallazgo analizado en *Transdigital y Mediatika*.
- **Hibridación y responsabilidad editorial:** La frontera difusa entre la intervención humana y la automática en procesos como la clonación de voz o la generación de vídeo (deepfakes) exige mecanismos de trazabilidad que hoy son insuficientes. La investigación subraya que, sin protocolos de supervisión (*human-in-the-loop*), la soberanía informativa se desplaza hacia infraestructuras globales ajenas al contexto vasco, como se argumenta en *Interconectando Saberes y Estudios sobre el Mensaje Periodístico*.
- **Iniciativas de corregulación:** Se identifican propuestas para mitigar estos vacíos, como la creación de marcos regulatorios alineados con la Ley de Inteligencia Artificial de la Unión Europea y el desarrollo de herramientas de verificación propias que devuelvan el control ético a las redacciones, según se extrae de los trabajos en *Mediatika* e *Interconectando Saberes*.

La investigación concluye que el modelo de gobernanza actual en el periodismo vasco es "defensivo". La cautela observada en las redacciones actúa como un sustituto temporal ante la falta de políticas corporativas sólidas. Sin embargo, esta situación de "indeterminación normativa" deja al periodista en una posición vulnerable: se le exige innovar, pero no se le proporcionan los protocolos para hacerlo de forma segura.

Por tanto, la sostenibilidad ética de la IA en Euskadi requiere pasar de la ética individual a la gobernanza institucional. La confianza de la audiencia no se garantiza mediante la exclusión de la tecnología, sino a través de la explicitación de sus límites. Esta tesis sostiene que la formalización de la auditoría algorítmica y la actualización de los códigos deontológicos son pilares imprescindibles para asegurar que la automatización no degrade la función democrática de los medios locales ni la soberanía sobre sus propios contenidos.

6.6. Gobernanza, soberanía tecnológica y desafíos en el contexto regional (OE6)

La investigación concluye que el ecosistema mediático vasco se encuentra en una situación de dependencia crítica respecto a las infraestructuras de las grandes corporaciones tecnológicas globales (*Big Tech*). Esta subordinación técnica plantea un riesgo real de "dilución territorial", donde las particularidades lingüísticas y culturales del País Vasco quedan desplazadas por lógicas algorítmicas estandarizadas. La respuesta del sector, aunque incipiente, apunta hacia la necesidad

de reclamar una soberanía tecnológica que permita desarrollar herramientas situadas y adaptadas a la realidad de una lengua minorizada y un entorno de proximidad.

Este análisis final se fundamenta en la integración de los hallazgos más estratégicos de todo el compendio:

- **Dependencia de infraestructuras globales:** El 92% de las herramientas de IA utilizadas por los periodistas vascos pertenecen a empresas externas (OpenAI, Google, Microsoft). Esta dependencia genera una vulnerabilidad económica y editorial, ya que los medios locales carecen de control sobre los sesgos o los cambios en los algoritmos de los que ahora dependen sus flujos de trabajo, tal como se desprende de los resultados en el artículo de la revista *Transdigital* y *Frontiers in Communication*.
- **Invisibilización del contexto local:** La experimentación técnica demuestra que los modelos de lenguaje globales tienen dificultades para reconocer la agenda informativa regional. En pruebas de actualidad vasca, solo el 15,5% de las fuentes citadas por la IA fueron medios del territorio, lo que confirma una tendencia a la centralización informativa que ignora la pluralidad mediática del País Vasco, según los datos del artículo en *Hipertext.net*.
- **El desafío lingüístico como eje de soberanía:** La eficacia de los modelos decae significativamente al operar en euskera en comparación con el castellano o el inglés. Esta brecha ha impulsado la creación de consorcios locales entre medios como *Berria* y centros de investigación para desarrollar modelos propios de traducción y síntesis de voz, entendiendo que el control del idioma en la máquina es la base de la soberanía digital, hallazgo detallado en *Estudios sobre el Mensaje Periodístico*.
- **Percepción de la IA como amenaza a la diversidad:** Un 48,2% de los profesionales encuestados teme que la adopción masiva de IA global homogeneice los contenidos y reduzca la riqueza de las narrativas locales. Este temor refuerza la tesis de que, en comunidades con identidad propia, la IA no puede ser un "producto empaquetado", sino que requiere una adaptación cultural profunda, tal como se analiza en *Transdigital* y *Mediatika*.
- **Hacia una gobernanza situada:** La investigación identifica que el éxito de la IA en Euskadi no vendrá de la cantidad de tecnología adoptada, sino de la capacidad de los medios para auditarla. La demanda de una "Gobernanza Ética Vasca" surge como una propuesta para que los medios del territorio establezcan sus propios estándares de calidad y equidad, protegiéndose de los sesgos de género y territoriales de los modelos estándar, según se extrae de *Interconectando Saberes* y *Mediatika*.

El análisis del OE6 permite concluir que el periodismo vasco no puede limitarse a ser un usuario pasivo de la IA global. Los resultados sugieren que nos encontramos ante

un desafío de supervivencia cultural en la era algorítmica. En este nivel estratégico, la 'cautela informativa' trasciende la rutina individual para convertirse en una postura política y de gestión. Los medios vascos han optado por una adopción fragmentada como medida de soberanía tecnológica. Esta cautela actúa aquí como un mecanismo de defensa institucional frente a la dependencia de las Big Tech, priorizando la protección del patrimonio lingüístico y la independencia del relato sobre la eficiencia algorítmica global.

Figura 135: Análisis de resultados



Fuente: Elaboración propia

Tabla 16: Tabla Resumen: Correspondencia de resultados, objetivos e hipótesis

Eje temático	Objetivo específico	Hipótesis validada	Resultado principal obtenido
Rutinas profesionales	OE1: Analizar dinámicas laborales	H1: Modificación de roles y tareas	Integración asistencial bajo el modelo de "cautela informativa".
Alfabetización	OE2: Evaluar percepciones y formación	H2: Mediación por falta de recursos	Brecha formativa crítica: solo un 14,1% posee formación reglada.
Multilingüismo	OE3: Eficacia en entornos bilingües	H3: Limitaciones de precisión en euskera	Opacidad territorial y dilución de la información local frente a la global.
Sesgos de género	OE4: Identificar sesgos algorítmicos	H4: Amplificación de desigualdades	Reproducción de estereotipos y visibilidad femenina reducida en outputs.
Transparencia	OE5: Examinar mecanismos de gobernanza	H5: Ausencia de protocolos formales	Accountability informal dependiente de la ética individual de cada medio.
Contexto regional	OE6: Contextualizar en políticas europeas	H6: Necesidad de gobernanza situada	Riesgo de dependencia estructural por asimetría de poder tecnológico.

Fuente: Elaboración propia

La revisión sistemática de los hallazgos permite confirmar la Hipótesis Principal de esta tesis: la incorporación de la IA en el País Vasco está transformando de manera estructural las prácticas periodísticas, si bien este proceso no es uniforme ni puramente tecnológico, sino que se desarrolla bajo tensiones profesionales profundas y riesgos socio-técnicos derivados de una alfabetización desigual.

Los resultados aquí presentados ofrecen una cartografía del cambio en un ecosistema periférico que, a pesar de su dimensión, muestra una notable capacidad de

resistencia ética. Esta resistencia se manifiesta a través de una cautela informativa que opera en dos niveles: como un protocolo de control de calidad en las rutinas diarias de los redactores (OE1) y como una reserva estratégica de soberanía frente a la dependencia de las infraestructuras globales (OE6).

La triangulación de datos demuestra que la IA no es un fenómeno neutro; es una tecnología que, sin una gobernanza situada basada en criterios de justicia de datos y equidad lingüística, compromete la pluralidad y la visibilidad de la realidad vasca. En última instancia, la tesis concluye que el futuro del periodismo en Euskadi dependerá de la capacidad institucional para cerrar el abismo formativo actual —fundamental para auditar sesgos de género y territoriales— y establecer protocolos de transparencia que devuelvan al periodista el control consciente y crítico sobre el discurso público en la era de la automatización.

7. Conclusiones y discusión

La presente investigación articula los hallazgos empíricos obtenidos en el ecosistema mediático vasco con el debate científico internacional, permitiendo una interpretación técnica y profesional de cómo la IA está reconfigurando la autoridad periodística. Al contrario de las visiones que plantean la automatización como un proceso uniforme, esta tesis demuestra que la integración de la IA es un fenómeno condicionado por la cultura profesional, las desigualdades estructurales del sector y la identidad territorial. La lectura de los seis artículos que componen este compendio permite superar la descripción de herramientas para profundizar en las dimensiones éticas y sociales del cambio. Se valida así una hipótesis principal que sostiene que, en contextos regionales bilingües, la transformación digital no es un proceso de sustitución, sino una negociación constante entre la búsqueda de eficiencia y la responsabilidad ineludible del control editorial humano.

7.1. La hibridación asistida y la "reserva de jurisdicción" profesional ante el algoritmo (OE1 - H1)

La primera cuestión fundamental que responde esta investigación es si la IA está desplazando la función del periodista o si, por el contrario, refuerza la necesidad de un criterio humano más estricto. La idea central que emana del análisis es que el periodismo vasco no vive un proceso de reemplazo, sino una reconfiguración donde el profesional delega tareas mecánicas, pero mantiene el control sobre el núcleo de la información mediante una "reserva de jurisdicción". Esta postura se apoya en los datos de los Artículos 1 y 4, que reflejan cómo la IA ha penetrado con éxito en procesos de soporte técnico como la transcripción automática (23%), la traducción (46,4%) o la corrección de estilo (27,6%), pero encuentra un límite claro en la generación de contenido informativo. El hecho de que solo el 13% de los medios tradicionales utilice esta tecnología para la redacción directa de noticias confirma que la adopción es inversamente proporcional a la complejidad del juicio ético requerido en cada pieza.

Este comportamiento de las redacciones vascas coincide con la noción de "apropiación sociotécnica" de Pinto y Barbosa (2024), donde se defiende que la tecnología es moldeada por el contexto que la recibe. Si bien autores como Diakopoulos (2019) plantean que la IA redefine las estructuras de rendición de cuentas, en el País Vasco esta transición se produce bajo un modelo de hibridación asistida: el periodista integra la herramienta como una extensión técnica que no suplanta su criterio. La aportación específica que el caso vasco ofrece al debate internacional es que esta resistencia a la automatización total no es un retraso, sino una medida de seguridad editorial. En un ecosistema donde los medios de proximidad son el referente de veracidad frente a la desinformación global, delegar la autoridad en algoritmos opacos se percibe como un riesgo para la reputación del medio. Por tanto, el caso vasco demuestra que, en comunidades con identidad propia, la "cautela" funciona como una metodología profesional para proteger la confianza de la audiencia y asegurar que la innovación no degrade la función social del periodismo.

7.2. El abismo formativo: de la precariedad técnica a la vulnerabilidad estratégica (OE2, OE6 - H2, H6)

La segunda cuestión central que se desprende de esta investigación es la contradicción estructural entre el interés profesional por la innovación y la parálisis formativa del sector. Esta tesis sostiene que el déficit de capacitación detectado en el ecosistema vasco no es un problema individual de los redactores, sino una vulnerabilidad estratégica que compromete la autonomía del sistema mediático en su conjunto. El hallazgo más crítico de esta investigación, desarrollado en el Artículo 3, es que apenas el 14,1% de los 501 periodistas encuestados posee formación reglada en IA. Esta carencia institucional obliga a que la inmensa mayoría de los profesionales recurra al autoaprendizaje informal, lo que genera una comprensión fragmentada y puramente instrumental de la tecnología. Los datos revelan, además, una preocupante asimetría: mientras que los perfiles técnicos presentan tasas de formación del 66,7%, los roles más cercanos a la toma de decisiones editoriales muestran niveles residuales, situándose en un 13,8% para redactores y un 6,3% en el caso de los presentadores.

Esta realidad académica y profesional en Euskadi dialoga con las advertencias de autores como Ng et al. (2021) y Sonni et al. (2024), quienes sostienen que la alfabetización mediática en la era de la IA debe integrar de forma indisoluble las dimensiones técnica, ética y crítica. Sin embargo, en el País Vasco, la ausencia de esta base formativa sistémica inhabilita al periodista para ejercer una supervisión consciente o el modelo de "humano en el bucle" (*human-in-the-loop*). Como indican Trillo-Domínguez y Alberich-Pascual (2020), la metamorfosis digital exige una explosión de creatividad, pero los resultados de esta tesis añaden que, sin una capacidad de auditoría técnica por parte de la redacción, esa innovación queda expuesta a alucinaciones algorítmicas y errores de bulto. El hecho de que un 63,91% de los profesionales identifique las barreras éticas como el principal freno para la adopción de la IA confirma que existe una voluntad de rigor, pero que esta carece de las herramientas técnicas para materializarse de forma segura.

La aportación específica del caso vasco a este debate internacional radica en demostrar que, en ecosistemas de escala limitada, el abismo formativo crea una dependencia ciega de las interfaces globales desarrolladas por las grandes corporaciones tecnológicas (*Big Tech*). Al no poseer conocimientos para auditar la "caja negra" del algoritmo, el periodista vasco queda reducido a un operario de software ajeno, perdiendo la soberanía sobre sus propios procesos de producción y verificación. Esta investigación aporta, por tanto, una evidencia necesaria para el marco de la Ley de Inteligencia Artificial de la Unión Europea: la transparencia algorítmica es inútil si el profesional que debe supervisarla no cuenta con la formación crítica para interpretar sus sesgos. En definitiva, esta tesis postula que la capacitación no es un complemento operativo, sino una condición política indispensable para

asegurar la supervivencia democrática y la independencia del periodismo de proximidad frente a la automatización global.

7.3. El desafío de la justicia lingüística y la "periferia digital" (OE3 - H3)

La tercera dimensión fundamental que aborda esta investigación se sitúa en la intersección entre la innovación tecnológica y la preservación de las lenguas minorizadas. La idea central es que los modelos de IA generativa actuales no son neutrales, sino que ejercen un "centralismo algorítmico" que degrada la eficacia informativa cuando se opera fuera de los marcos lingüísticos y geográficos dominantes. Esta tesis sostiene que, para un ecosistema mediático bilingüe, la adopción de herramientas globales sin una adaptación local previa no supone un avance, sino un riesgo de invisibilización cultural y pérdida de precisión periodística.

Esta afirmación se sustenta en los resultados experimentales detallados en el Artículo 2, donde se realizó una auditoría comparada entre cinco de los principales modelos de lenguaje (ChatGPT, Claude, Gemini, Copilot y Perplexity). La evidencia es concluyente: existe una brecha de rendimiento significativa cuando se consulta sobre la actualidad vasca en euskera en comparación con el castellano o el inglés. Los datos muestran que la tasa de "alucinaciones" y la falta de trazabilidad documental aumentan drásticamente en la escala regional; de las fuentes citadas por estos modelos en casos de proximidad, solo el 15,5% correspondían a medios de comunicación del territorio (como Eitb, Berria o Deia), mientras que el 67,4% de las referencias eran de ámbito nacional. Este hallazgo valida la Hipótesis 3, demostrando que la opacidad territorial de los algoritmos actuales diluye la información de cercanía en favor de relatos centralizados.

Desde el plano de la literatura científica, estos resultados dialogan con las advertencias de Tomassi et al. (2024) y Bontridder & Pouillet (2021) sobre la "dilución del contexto" en la era de la automatización. Si bien la IA promete una mayor eficiencia en la producción de contenidos, esta investigación demuestra que esa eficiencia tiene un coste en términos de pluralismo territorial. Al priorizar datasets masivos en lenguas mayoritarias, los modelos de lenguaje comerciales actúan como una fuerza de homogeneización cultural, lo que autores como Young (2020) definen como una forma de injusticia epistémica. La dependencia de infraestructuras anglocéntricas o hispanocéntricas constituye, por tanto, una barrera para que el periodismo vasco pueda ejercer su función de servicio público con garantías de rigor y representatividad.

La aportación específica del caso vasco a este debate internacional es la constatación de que la soberanía lingüística es, en el siglo XXI, una cuestión de soberanía tecnológica. Los resultados de esta tesis revelan que los medios vascos (como el consorcio liderado por Berria o los proyectos de Eitb) no están buscando

simplemente "usar" la IA, sino que están impulsando el desarrollo de infraestructuras propias —traductores neuronales y sintetizadores de voz específicos— para mitigar el sesgo de las plataformas globales. El caso vasco aporta una lección crítica para otras comunidades lingüísticas: en la periferia digital, la supervivencia del periodismo de proximidad depende de la capacidad de reclamar una "agencia algorítmica" que permita auditar y corregir el código para que este reconozca y respete la diversidad cultural del territorio.

7.4. El algoritmo como espejo: sesgos de género e integridad del discurso informativo (OE4 - H4)

La cuarta dimensión crítica de esta investigación aborda la capacidad de la IA para actuar como un agente de reproducción y amplificación de desigualdades históricas. La idea central es que la IA informativa no opera en un vacío de neutralidad técnica, sino que funciona como un "espejo deformante" que cristaliza los sesgos de género presentes en los *datasets* de entrenamiento. Esta tesis sostiene que la delegación de la producción periodística en modelos globales no auditados supone un riesgo directo para el pluralismo y la equidad, especialmente en ecosistemas que, como el vasco, han realizado esfuerzos significativos por alcanzar la paridad en la esfera pública.

Los hallazgos empíricos desarrollados en el Artículo 6 confirman esta tendencia mediante un análisis cualitativo y documental de los *outputs* algorítmicos. La evidencia revela una "segregación temática" sistémica: mientras los sistemas de IA priorizan fuentes masculinas para la cobertura de áreas de autoridad como la economía o la tecnología, las representaciones femeninas se concentran de forma desproporcionada en ámbitos vinculados al entretenimiento o los cuidados. A esto se suma un patrón de diseño detectado en la feminización de las interfaces de asistencia técnica, que refuerza roles de servidumbre. Estos datos validan la Hipótesis 4, demostrando que la automatización tiende a invisibilizar a las mujeres en roles de liderazgo informativo, lo que compromete la integridad del discurso periodístico y la justicia epistémica en el territorio.

Desde el plano de la literatura científica, estos resultados dialogan con los marcos de justicia de datos de Fricker (2007) y Showkat (2022), quienes advierten que la desinformación de género y los sesgos algorítmicos no son fallos accidentales, sino características intrínsecas de modelos entrenados con datos históricos desiguales. Si bien autores como Noble (2018) han denunciado los "algoritmos de opresión", esta investigación traslada el debate al ámbito de las redacciones bilingües, donde la intersección entre género y lengua minorizada añade una capa extra de vulnerabilidad. La falta de equipos de desarrollo diversos y la opacidad de los modelos de lenguaje comerciales (LLMs) dificultan la detección de estas discriminaciones antes de que el contenido llegue a la audiencia, lo que exige una gobernanza proactiva que trascienda la simple eficiencia operativa.

La aportación específica del caso vasco a este debate internacional radica en la exigencia de una "gobernanza de género situada". Los resultados demuestran que, ante la incapacidad de los modelos estándar para corregir sus propios sesgos, los medios vascos deben integrar una alfabetización crítica que capacite a los periodistas para auditar el código desde una perspectiva interseccional. El caso vasco aporta una lección fundamental: la lucha por la igualdad en el periodismo del siglo XXI ya no se libra solo en los organigramas de las redacciones, sino en la supervisión ética de los datos. Esta tesis concluye que la verdadera innovación en un ecosistema democrático exige recuperar el control humano sobre la "caja negra" algorítmica para asegurar que la automatización no degrade los avances sociales logrados en materia de representación y equidad.

Tras el análisis integral de las evidencias obtenidas en el compendio de seis artículos que conforman esta investigación, se concluye que la integración de la IA en el periodismo no es un proceso tecnológico universal, sino un fenómeno profundamente situado y condicionado por el ecosistema receptor. El caso vasco atestigua que el contexto —entendido como la intersección entre una identidad cultural propia, una realidad bilingüe y una escala mediática de proximidad— actúa como un filtro determinante que condiciona el modo, el ritmo y la profundidad de la apropiación técnica. Esta investigación demuestra que, frente a la presión por una eficiencia de mercado impuesta por las grandes corporaciones tecnológicas, el ecosistema mediático vasco ha optado por una transición marcada por la asimetría y la cautela, donde la innovación se subordina a la preservación del contrato de confianza con la audiencia.

Esta postura se manifiesta en una conclusión que se considera central en este trabajo: la identificación de una "cautela informativa" que recorre transversalmente todo el ecosistema. Basándose en los resultados obtenidos, se sostiene que esta cautela no debe interpretarse como un retraso técnico o un síntoma de resistencia al progreso, sino como una estrategia de soberanía editorial fundamentada en la responsabilidad profesional. Los periodistas vascos han decidido actuar como curadores algorítmicos, reclamando un control humano estricto (*human-in-the-loop*) tras constatar que las herramientas globales no responden con precisión a su realidad. Esta desconfianza está plenamente justificada por los hallazgos de la experimentación técnica, que determinan la existencia de un sesgo territorial sistémico en la IA generativa actual. Los grandes modelos de lenguaje invisibilizan la actualidad regional, omiten fuentes de proximidad —con apenas un 15,5% de representación local— y presentan una degradación severa de la calidad informativa cuando operan en euskera. Por tanto, la automatización global, si no es intervenida mediante una adaptación cultural profunda, actúa como un vector de homogeneización que erosiona el pluralismo territorial y subordina el relato local a lógicas centralistas ajenas a la idiosincrasia del territorio, lo que constituye una forma de colonialismo digital.

Asimismo, se concluye que el mayor riesgo para la integridad del periodismo vasco no reside en la tecnología por sí misma, sino en la alarmante brecha entre la velocidad de implementación y la capacidad de auditoría profesional. El hecho de que solo un 14,1% de los periodistas posea formación reglada constituye un abismo formativo que inhabilita al profesional para detectar sesgos de género y territoriales inscritos en el código algorítmico. Se determina, por tanto, que esta carencia, sumada a la ausencia de protocolos institucionales de gobernanza, deja la responsabilidad ética en manos de la voluntad individual de cada redactor, generando una zona de indeterminación normativa que debilita la autoridad periodística ante la desinformación. Ante este panorama, se plantea que el futuro del sector exige transitar de la autorregulación informal a una responsabilidad institucional proactiva, donde los medios implementen guías de estilo y protocolos de transparencia que garanticen la trazabilidad del contenido automatizado.

Esta tesis postula, en última instancia, que el reto de la IA en el País Vasco es un desafío democrático que requiere la creación de una IA situada o de proximidad. Esto implica que la formación debe dejar de ser un barniz operativo para convertirse en una prioridad política, capacitando a los periodistas no solo para usar la tecnología, sino para auditarla y neutralizar sus discriminaciones. La supervivencia del periodismo de proximidad dependerá de fomentar la colaboración para desarrollar infraestructuras tecnológicas soberanas que dominen el euskera y reconozcan el contexto sociopolítico vasco. En definitiva, la agencia humana, la supervisión ética y el compromiso innegociable con la diversidad lingüística deben ser los pilares sobre los que se construya la verdad pública en la era de la automatización. De este modo, el éxito de esta transición no vendrá determinado por la potencia de cálculo de las máquinas, sino por la capacidad de los profesionales para reclamar y ejercer una soberanía algorítmica propia.

9. Limitaciones del trabajo

Aunque la tesis aporta un conocimiento sólido y situado sobre el impacto de la IA en el periodismo vasco, es necesario reconocer un conjunto de limitaciones que permiten contextualizar el alcance de los resultados y abren nuevas líneas de investigación. Estas limitaciones no cuestionan la validez del trabajo, pero sí ayudan a entender mejor qué puede afirmarse con contundencia y qué debe leerse con cautela.

La primera limitación deriva de la propia naturaleza territorial del estudio. El ecosistema mediático del País Vasco —marcado por la cooficialidad lingüística, la estructura institucional autonómica, la presencia de un servicio público robusto y la centralidad cultural del euskera— configura un escenario singular para la adopción de IA. Esta especificidad es, al mismo tiempo, una fortaleza y un límite: permite profundizar en un caso especialmente sensible a la diversidad lingüística, pero dificulta extrapolar de forma directa los resultados a contextos monolingües o altamente centralizados. Aunque algunos patrones identificados (cautela editorial, brecha formativa, riesgo de sesgos, gobernanza débil) pueden aparecer en otros territorios periféricos o multilingües, sería inapropiado universalizarlos sin atender a sus particularidades institucionales y socioculturales.

Una segunda limitación procede del acceso restringido a información técnica sobre los sistemas de IA empleados por los medios y plataformas digitales. La investigación no ha podido examinar en profundidad elementos clave como los datasets de entrenamiento, los criterios de optimización de los modelos o las auditorías internas aplicadas a herramientas generativas utilizadas en redacciones. Esta opacidad —habitual en entornos institucionales y comerciales— coincide con uno de los problemas que la propia tesis identifica: la falta de trazabilidad y rendición de cuentas en los procesos automatizados. La imposibilidad de acceder a estos componentes técnicos obliga a centrarse en los usos, percepciones y efectos observables, más que en los mecanismos internos de los sistemas algorítmicos.

A ello se suma una limitación vinculada al ritmo vertiginoso de la innovación tecnológica. Los modelos generativos evolucionan con tal rapidez que algunas herramientas analizadas durante la investigación —como ChatGPT, Gemini o Synthesia— experimentaron cambios sustantivos mientras se redactaban los artículos. Esta velocidad introduce un desfase inevitable entre la realidad tecnológica y la capacidad de la investigación para captarla en tiempo real. Lejos de ser un obstáculo propio de este estudio, representa un desafío estructural para toda investigación sobre IA, que exige metodologías flexibles y actualizaciones constantes para seguir el pulso de la transformación digital.

Algunos de los conceptos clave movilizados en esta tesis —transparencia algorítmica, ética de la IA, alfabetización mediática, justicia epistémica— siguen siendo categorías abiertas en la literatura y no están plenamente consensuadas. Aunque se ha llevado a cabo un esfuerzo riguroso para definir las y aplicarlas de manera coherente, su operacionalización empírica puede variar según las tradiciones disciplinarias y los

enfoques teóricos. Esta variabilidad no reduce su utilidad analítica, pero sí invita a interpretar los resultados dentro del marco conceptual específico adoptado en la tesis.

Es por todo ello que estas limitaciones no restan fuerza a las conclusiones, pero sí subrayan la importancia de continuar investigando: ampliando el análisis a otros territorios, exigiendo mayor transparencia a las instituciones tecnológicas, observando la evolución acelerada de los modelos generativos y afinando marcos conceptuales que permitan comprender con precisión las implicaciones democráticas de la IA en los ecosistemas mediáticos contemporáneos.

10. Propuestas para futuras líneas de investigación

Los hallazgos obtenidos en esta tesis, así como las limitaciones identificadas, permiten delinear una agenda de investigación futura centrada en el análisis crítico, situado y plural de la IA en los medios de comunicación. A continuación, se proponen varias líneas prioritarias que pueden contribuir a una comprensión más profunda, comparativa y propositiva del fenómeno.

En primer lugar, resulta urgente desarrollar estudios longitudinales que permitan observar cómo evoluciona la integración de la IA en las rutinas periodísticas, en la organización del trabajo y en la cultura profesional de los medios. Hasta ahora, la mayoría de los estudios —incluidos los aquí presentados— ofrecen una fotografía estática del momento actual. Sin embargo, procesos como la normalización del uso de herramientas generativas, la transformación del rol del periodista o la adaptación organizativa requieren ser comprendidos en términos de cambio progresivo. Estos estudios longitudinales permitirían detectar no solo tendencias de adopción o rechazo, sino también resistencias simbólicas, formas de apropiación creativa y redefiniciones normativas a lo largo del tiempo.

Otra línea de investigación relevante se refiere a la gobernanza tecnológica dentro de las organizaciones periodísticas. Aunque algunos artículos han abordado cuestiones como la transparencia o la responsabilidad editorial, todavía se requiere un análisis más profundo sobre los procesos de toma de decisiones tecnológicas en las redacciones: ¿quién decide qué herramienta se incorpora?, ¿bajo qué criterios?, ¿existe participación profesional o es una decisión puramente técnica?, ¿cómo se gestiona la tensión entre innovación y control editorial? Estudios etnográficos o de observación participante podrían iluminar estas dinámicas y contribuir a una visión más realista sobre las condiciones institucionales de uso de la IA.

Asimismo, se propone explorar modelos de diseño participativo de tecnologías de IA para el periodismo. Frente a un panorama dominado por herramientas creadas por grandes plataformas, es necesario fomentar desarrollos tecnológicos que respondan a las necesidades reales de periodistas y comunidades. Simultáneamente, el trabajo colaborativo entre desarrolladores, profesionales de la información, académicos y representantes de la sociedad civil podría dar lugar a soluciones más éticas, inclusivas y contextualizadas. Este enfoque, basado en los principios de la co-creación y la innovación abierta, representa una oportunidad para democratizar el desarrollo tecnológico y reducir las asimetrías de poder actualmente existentes.

Otra línea prioritaria concierne al análisis sistemático de los sesgos algorítmicos, no solo desde una perspectiva de género —como ya se ha abordado en esta tesis—, sino también considerando otras dimensiones como la etnicidad, la orientación sexual, la edad, la clase social o el territorio. En un contexto en el que los algoritmos participan en la selección, jerarquización y distribución de contenidos, comprender cómo se configuran y perpetúan determinadas exclusiones es una tarea urgente. Ello implica, entre otras cosas, auditar los datasets, revisar los criterios de entrenamiento,

fomentar la diversidad en los equipos de desarrollo y establecer mecanismos independientes de supervisión y corrección de sesgos.

Finalmente, se plantea como línea de investigación estratégica el desarrollo de estudios comparados entre regiones, estados o sistemas mediáticos. Si bien esta tesis ha apostado por una lectura contextualizada de la automatización, enraizada en el contexto vasco, la comparación con otras realidades —especialmente aquellas con características similares, como territorios multilingües, regiones periféricas o medios locales con escasa digitalización— podría ofrecer aprendizajes relevantes. Estos estudios permitirían identificar patrones comunes, divergencias culturales, marcos regulatorios contrastantes o buenas prácticas transferibles. Además, contribuirían a enriquecer los debates internacionales sobre la ética de la IA, la soberanía tecnológica y el futuro del periodismo.

11. Anexos

11.1.Trabajos publicados



La inteligencia artificial en el periodismo: impacto social, ética y transparencia en el contexto del País Vasco

Artificial Intelligence in Journalism: Social Impact, Ethics, and Transparency in the Basque Country Context

Barbara Sarrionandia Urquien *

Universidad del País Vasco / Euskal Herriko Unibertsitatea

bsarrionandia001@ikasle.ehu.eus

ORCID: <https://orcid.org/0009-0006-0053-3730>

Sección: Artículo de Investigación.

Fecha de recepción:

Fecha de aceptación:

La inteligencia artificial en el periodismo: impacto social, ética y transparencia en el contexto del País Vasco

Artificial Intelligence in Journalism: Social Impact, Ethics, and Transparency in the Basque Country Context

Resumen

La implantación de la Inteligencia artificial (IA) en el periodismo plantea importantes desafíos y oportunidades, transformando tanto las prácticas profesionales como los principios éticos del sector. Este artículo analiza el impacto de la IA en los medios de comunicación del País Vasco mediante un enfoque metodológico mixto que combina encuestas a 501 periodistas en activo y entrevistas en profundidad con responsables de tecnología e innovación de seis importantes medios de la región. Los resultados muestran que el 45,91% de los encuestados nunca utiliza IA en su labor profesional, y un 63,91% cita preocupaciones éticas como principal barrera. Además, se detecta una falta de formación específica, lo que evidencia la desconexión entre el potencial tecnológico y las competencias actuales del sector. Este estudio destaca la necesidad de establecer normativas claras, fomentar la transparencia y diseñar programas de formación continua para garantizar un uso ético y efectivo de esta tecnología.

Palabras clave: Inteligencia artificial, periodismo, ética, desinformación, transparencia informativa.

Abstract

The implementation of artificial intelligence (AI) in journalism presents significant challenges and opportunities, transforming both professional practices and the ethical principles of the sector. This article analyzes the impact of AI on the media landscape in the Basque Country through a mixed-methods approach that combines surveys of 501 active journalists with in-depth interviews conducted with heads of technology and innovation from six major media outlets in the region. The results show that 45.91% of respondents never use AI in their professional work, and 63.91% cite ethical concerns as the main barrier. Additionally, a lack of specific training is evident, highlighting the disconnect between technological potential and the sector's current skill set. This study underscores the need to establish clear regulations, promote transparency, and design continuous training programs to ensure the ethical and effective use of this technology.

Keywords: Artificial intelligence, journalism, ethics, disinformation, informational transparency.

1. Introducción

La inteligencia artificial (IA) ha emergido como una tecnología transformadora con impacto en múltiples sectores, incluido el periodismo. Su integración está redefiniendo la forma en que se recopilan, procesan y distribuyen las noticias. Sin embargo, este proceso implica tanto oportunidades como desafíos. Los periodistas deben adaptarse a estas herramientas tecnológicas mientras conservan la integridad de su labor en un contexto de creciente desinformación y postverdad (Salaudeen, 2022; Ramon, Mauri-Ríos & Díaz-Campo, 2020; Ballesteros-Aguayo et al., 2022).

En esta línea, las aplicaciones de la IA en el periodismo incluyen el análisis de datos, la generación automatizada de contenidos y la personalización de noticias. Estas herramientas ya están alterando estructuras empresariales en medios de comunicación (Sánchez-García et al., 2023), especialmente en contextos como el País Vasco. A nivel internacional, el Reuters Institute destaca el avance de tecnologías como el aprendizaje automático y la generación de lenguaje natural, con potencial para transformar la producción de contenidos (Newman, 2020).

Desde una perspectiva conceptual, la IA se entiende como el uso de sistemas computacionales que procesan datos y simulan razonamientos humanos (Giletta et al., 2020). Su expansión en el ámbito periodístico responde a avances en capacidad de cómputo y acceso a grandes volúmenes de datos (Túñez-López, Feiras-Ceide & Vaz-Álvarez, 2021). No obstante, esta adopción genera un debate entre sus beneficios y los retos que plantea, como la automatización de tareas o el riesgo de pérdida de control editorial (Diakopoulos, 2020; Parratt-Fernández, Mayoral-Sánchez & Mera-Fernández, 2021).

Entre los beneficios, se destaca la automatización de tareas repetitivas, la personalización de contenidos y la verificación automatizada de información (Noain-Sánchez, 2022; Ufarte-Ruiz, Calvo Rubio & Murcia Verdú, 2021). Estas aplicaciones permiten mejorar la eficiencia en la redacción, optimizar recursos y combatir fenómenos como la desinformación. Sin embargo, la opacidad de los algoritmos, la reproducción de sesgos y la falta de rendición de cuentas generan inquietudes sobre su impacto en la transparencia editorial (Porlezza, 2023; Helberger & Diakopoulos, 2023).

La literatura científica insiste en la necesidad de una supervisión humana constante. Esta supervisión es clave para mantener el rol crítico del periodismo y su compromiso con la verdad (Deuze & Beckett, 2022; Kothari & Hickerson, 2020). Asimismo, se plantea la urgencia de ofrecer formación específica en IA a los profesionales de los medios, que incluya tanto habilidades técnicas como competencias éticas (Túñez-López et al., 2021). La falta de formación puede derivar en una dependencia acrítica de tecnologías que no siempre garantizan equidad ni transparencia.

En cuanto al marco normativo, se advierte una carencia de regulaciones específicas para el uso de IA en medios. Esta ausencia normativa puede afectar la confianza del público y perpetuar desigualdades (Porlezza, 2023; Helberger & Diakopoulos, 2023). Establecer directrices claras sobre responsabilidades y transparencia es fundamental para fortalecer el rol social del periodismo y garantizar el derecho ciudadano a una información veraz.

La IA en el periodismo plantea así una dualidad: ofrece eficiencia y nuevas posibilidades narrativas, pero también desafíos éticos y sociales que deben ser atendidos críticamente. Esta investigación, por tanto, parte de

un enfoque empírico para analizar la implementación de estas tecnologías en el contexto vasco, mediante encuestas y entrevistas con profesionales de la comunicación.

1.1 Tensiones y oportunidades en la aplicación de IA en el periodismo actual

La IA ha revolucionado numerosos sectores, y su aplicación en los medios ha generado una de las transformaciones más profundas en la forma de producir y consumir información. Definida como tecnología capaz de procesar grandes volúmenes de datos simulando razonamiento humano (de Lima-Santos & Ceron, 2022; Benbya, Davenport & Pachidi, 2020), su papel en los medios se asocia tanto con la innovación como con la necesidad de encontrar modelos de negocio sostenibles (Peña-Fernández et al., 2023). Uno de los desarrollos más significativos es el de los modelos de lenguaje generativo, como ChatGPT 3.5 y 4 de OpenAI, que han mostrado gran capacidad para generar textos similares al lenguaje humano (Pavlik, 2023; Hudíková, 2023). Empresas como Anyword, Jasper o Copy.ai han adaptado estos sistemas a tareas de redacción automatizada en diversos campos, incluido el periodismo (Biswas, 2023; Lopezosa et al., 2023; Jones, Luger & Jones, 2023).

Estos avances permiten optimizar tiempos de producción y personalizar contenidos, aunque también ofrecen herramientas de búsqueda, transcripción y etiquetado automatizado (Subiela-Hernández & Vizcaino-Laorga, 2023; Espinosa, Alsarayreh & Benítez, 2024). Asimismo, su contribución a la lucha contra la desinformación mediante verificación automatizada ha sido destacada por varios estudios (Canavilhas, 2022).

No obstante, estos beneficios conviven con riesgos. La priorización de contenidos sensacionalistas por parte de los algoritmos y la creación de noticias falsas a través de modelos generativos amenazan la credibilidad periodística (Gutiérrez López et al., 2023; Opdahl et al., 2023; Leiser, 2022). El informe de Newsguard (2023) documenta más de 450 sitios web creados casi íntegramente por IA, sin supervisión editorial y con contenido engañoso.

Otro punto de fricción es la precarización del empleo. La automatización de funciones antes humanas genera incertidumbre en las redacciones, mientras que la concentración de datos en grandes plataformas tecnológicas limita la diversidad informativa (Dhiman, 2023; Salaverria, 2023; Franganillo, 2023). Esto afecta tanto a periodistas como a consumidores, que dependen de un ecosistema plural para una información de calidad.

La ética es otro eje central. La opacidad algorítmica y la falta de transparencia sobre el funcionamiento de estas tecnologías preocupan a investigadores y profesionales. El público debe ser informado de cuándo y cómo se emplea IA en la creación de contenidos (Ufarte-Ruiz, Calvo Rubio & Murcia-Verdú, 2021; Sanahuja & Esteban, 2023). Además, la protección de los datos usados en el entrenamiento de modelos también requiere atención (Vargues-Vara, 2024).

Aunque históricamente se ha prestado poca atención a la dimensión ética, actualmente cerca de una cuarta parte de los estudios sobre IA en periodismo abordan estas implicaciones (Lassi, 2022; Ventura, 2023). Existe consenso sobre la necesidad de regulaciones claras que asignen responsabilidades y limiten los riesgos (Porlezza, 2023; Helberger & Diakopoulos, 2023).

En Europa, marcos como la Ley de Servicios Digitales (DSA) y la Ley de Mercados Digitales buscan garantizar un uso seguro de estas tecnologías (Díaz & Jiménez, 2022). No obstante, deben ajustarse al uso periodístico específico, lo que exige coordinación entre actores públicos y privados (Fernández Torralvo, 2023;

Ventura-Salom, 2024). En España, aunque el código ético de la FAPE aún no contempla la IA, se han sugerido recomendaciones en esta línea (Peña-Fernández et al., 2023).

En el País Vasco, en concreto, el Colegio de Periodistas ha promovido directrices específicas sobre el uso ético de la IA, elaboradas junto a la Diócesis de Bilbao. Estas incluyen principios como la transparencia, la ciberseguridad y la formación continua (Bkareaga, 2024). Además, el Centro Vasco de Inteligencia Artificial (BAIC) impulsa una adopción responsable de estas herramientas en la región (BAIC, s.f.).

Se puede decir por tanto que el uso de IA en el periodismo plantea una dualidad crítica. Si bien ofrece oportunidades de mejora en eficiencia y acceso, también introduce riesgos que requieren regulación, ética y formación profesional constante. Solo mediante un enfoque integral será posible aprovechar su potencial sin comprometer los valores fundamentales del periodismo.

2. Método de investigación

La presente investigación se centra en analizar el impacto social y ético del uso de la inteligencia artificial (IA) en el periodismo del País Vasco, empleando una metodología mixta que combina enfoques cuantitativos y cualitativos. Este enfoque se justifica por la necesidad de captar una comprensión integral de las percepciones, prácticas y desafíos asociados a la implementación de la IA en el ámbito periodístico. La muestra incluye tanto a periodistas en activo como a responsables de áreas de innovación y tecnología de los principales medios de comunicación de la región.

El componente cuantitativo de la investigación se basa en una encuesta online administrada durante los meses de mayo y junio de 2024, dirigida a una muestra de 501 periodistas en activo. Los participantes fueron seleccionados a partir de la Guía de Comunicación Abierta del Gobierno Vasco (<https://gida.irekia.euskadi.eus>), que proporciona una base de datos completa de medios y profesionales que operan en Euskadi, y miembros de la Asociación Vasca de Periodistas. La encuesta, diseñada para garantizar el anonimato y la confidencialidad de los encuestados, incluyó preguntas relacionadas con la percepción de los riesgos y beneficios de la IA, la existencia de normativas específicas en sus organizaciones, la transparencia en el uso de estas herramientas y las preocupaciones éticas asociadas.

Los datos recopilados reflejan una amplia diversidad demográfica, incluyendo 276 hombres (55,1%), 223 mujeres (44,5%) y 2 participantes que se identificaron en la categoría de "otros" (0,4%). Además, se tuvieron en cuenta variables como los años de experiencia profesional, el tipo de medio en el que trabajan, el ámbito de difusión del mismo y si ocupan o no un puesto de responsabilidad. Esta diversidad permitió capturar un panorama representativo del ecosistema mediático en el País Vasco.

Complementariamente, el componente cualitativo de la investigación consistió en seis entrevistas en profundidad con responsables de áreas relacionadas con tecnología, innovación e IA en medios de comunicación

relevantes de la región. Estas entrevistas, realizadas entre abril y mayo de 2024, tuvieron una duración de entre 30 y 50 minutos y se llevaron a cabo tanto en formato presencial como mediante videoconferencia, dependiendo de la disponibilidad de los participantes. Las conversaciones abordaron cuestiones clave, como la implementación de políticas relacionadas con la IA, los desafíos específicos percibidos en el uso de estas herramientas y las estrategias adoptadas para mitigar riesgos éticos y sociales.

Tabla 1
Entrevistas en profundidad

Código	Nombre y apellidos	Medio	Tipo	Cargo
I1	Iván Pérez	Vocento /El Correo	Prensa	Responsable de nuevas tendencias tecnológicas y de calidad
I2	Jon Tellexea	EITB	Radiotelevisión pública	Subdirector de EITB Media
I3	Javier Turrado	Grupo Noticias/Deia	Prensa	Dirección del área tecnológica
I4	Beñat Zaldúa	Naiz	Prensa	Responsable de contenidos y previsiones
I5	Pello Urcelai	Berria	Prensa	Responsable de tecnologías de la información
I6	Jon Egaña	Cadena Ser / Radio Bilbao	Radio	Jefe de redacción

A partir de estos dos métodos, el estudio busca responder a las siguientes preguntas de investigación:

1. ¿Qué percepción tienen los periodistas del País Vasco sobre la utilidad y los riesgos asociados a la IA?
2. ¿Existen normativas específicas en los medios de comunicación para regular el uso de la IA?
3. ¿De qué manera los responsables de medios integran estrategias éticas para el uso de herramientas de IA?

4. ¿Qué preocupaciones éticas y sociales se identifican como prioritarias en la adopción de estas tecnologías en el ámbito periodístico?

La combinación de métodos cuantitativos y cualitativos permite no solo cuantificar tendencias y patrones en las percepciones de los profesionales, sino también contextualizar estos resultados mediante las experiencias y perspectivas concretas de los responsables de los medios. Este enfoque garantiza una triangulación de datos que refuerza la validez y profundidad de los hallazgos. Además, la inclusión de variables como el género, los años de experiencia y el tipo de medio amplía el espectro de análisis, proporcionando un marco más sólido para interpretar los resultados.

En este sentido, esta metodología permite abordar de manera integral los retos, oportunidades y dilemas éticos que plantea la inteligencia artificial en el periodismo, aportando una visión crítica y reflexiva sobre su impacto en el ecosistema mediático del País Vasco.

3. Resultados

El uso de la inteligencia artificial (IA) en el periodismo del País Vasco mostró un panorama caracterizado por su limitado alcance, pero con señales de interés emergente entre ciertos sectores. Los datos recopilados en la encuesta reflejaron que un 45,91% de los periodistas no utiliza nunca herramientas de IA en su trabajo diario, mientras que solo un 7,19% las emplea con regularidad diaria. Estas cifras destacaron la baja integración de la IA en las rutinas periodísticas, una situación que podía estar asociada a múltiples factores estructurales, culturales y de formación.

La reticencia al uso de la IA se explicaba, en parte, por barreras organizativas y culturales. En los medios tradicionales, como la prensa escrita y la televisión, persistía un enfoque conservador hacia la producción informativa, basado en prácticas manuales y un rechazo generalizado a la automatización. Esto contrastaba con los medios digitales, donde la tecnología formaba parte intrínseca de las dinámicas de trabajo, permitiendo una integración más fluida de herramientas basadas en IA.

Otro factor clave era la falta de formación técnica entre los profesionales de la comunicación. Los periodistas con más de 20 años de experiencia profesional, en particular, enfrentaban dificultades para adaptarse a las nuevas herramientas, lo que reforzaba la brecha generacional en la adopción de la IA. Por el contrario, los periodistas con menos de cinco años en la profesión mostraban una actitud más receptiva hacia estas tecnologías, lo que sugería que las nuevas generaciones podrían liderar el cambio en el sector.

Las diferencias en la adopción de la IA también se evidenciaban al analizar el tipo de medio en el que trabajan los encuestados. Los periodistas que emplean la IA de manera más frecuente suelen desempeñarse en

medios digitales, donde la tecnología estaba más integrada en los procesos de producción. Esto podía atribuirse a la necesidad de gestionar grandes volúmenes de información y responder rápidamente a las demandas del público.

Por otro lado, los medios tradicionales, como la prensa escrita, la televisión y la radio, presentan una adopción significativamente menor de estas tecnologías. Esto podía explicarse por la estructura jerárquica de estos medios, que tiende a priorizar las metodologías de trabajo convencionales. Además, los recursos limitados en términos de infraestructura tecnológica y personal capacitado agravaban esta situación, especialmente en los medios locales y regionales.

Tabla 2
Frecuencia de uso de la IA según tipo de medio

Frecuencia de uso	Prensa escrita (%)	Radio (%)	Televisión (%)	Digital (%)
Nunca	54,20	32,10	40,00	10,50
Alguna vez al mes	30,00	41,00	35,00	25,00
Semanalmente	10,00	21,50	18,00	50,00
Diariamente	5,80	5,40	7,00	14,50

Pese a las limitaciones actuales, el interés emergente por la IA es una señal alentadora. El grupo de adopción temprana, compuesto por el 20,56% de periodistas que utilizaban la IA semanalmente, podría desempeñar un papel clave en la promoción de estas herramientas dentro del sector. Este grupo, predominantemente asociado a medios digitales y funciones innovadoras, encontró en la IA una herramienta para optimizar tareas como la transcripción de entrevistas, la personalización de contenidos y la identificación de tendencias en redes sociales.

3.1. Preocupaciones éticas como barrera

La falta de adopción de la inteligencia artificial (IA) en el periodismo no solo se debía a la escasez de formación técnica o recursos tecnológicos, sino también a las preocupaciones éticas que rodean su

Hombres	52,38
Mujeres	46,94
Más de 20 años	47,62
Entre 10 y 20 años	19,05
Entre 5 y 10 años	17,69
Menos de 5 años	15,65

3.2. Transparencia y comunicación con las audiencias

La transparencia emergió como uno de los pilares fundamentales en la relación entre la inteligencia artificial (IA) y el periodismo. Los resultados de las encuestas mostraron que un 55,09% de los periodistas considera que es imprescindible comunicar explícitamente el uso de IA en cualquier etapa del proceso informativo, mientras que un 33,13% cree que solo debe notificarse cuando la IA actúa como fuente principal. En contraste, un 6,39% considera que no es necesario informar sobre el uso de esta tecnología.

La percepción de transparencia no solo afecta a las audiencias, sino que también influye en la confianza en los medios de comunicación. Los periodistas entrevistados destacaron que la falta de claridad sobre el uso de IA puede erosionar la confianza del público, especialmente si surgen casos de desinformación o uso indebido de estas tecnologías. Por otro lado, la comunicación proactiva sobre el uso de IA puede posicionar a los medios como líderes en innovación y ética, reforzando su credibilidad.

Aunque existe un consenso general sobre la relevancia de la transparencia, los periodistas señalaron varios obstáculos para su aplicación, como la ausencia de normativas claras, la falta de conocimientos técnicos suficientes para explicar el uso de la IA y el temor a que una comunicación demasiado especializada genere rechazo en el público.

3.3. Percepción de la desinformación y sesgos digitales

La relación entre la inteligencia artificial (IA) y los riesgos asociados a la desinformación y los sesgos digitales constituyó una de las principales preocupaciones de los periodistas encuestados. Los resultados indicaron que un 55,89% de los participantes consideraron que la IA incrementará significativamente los riesgos de desinformación, mientras que un 33,93% creyó que el impacto será moderado. Solo un 9,18% percibió un impacto limitado, y un 1% creyó que no tendrá ningún efecto.

La preocupación por la desinformación está estrechamente vinculada al potencial de la IA para generar contenidos falsos o manipulados, como los deepfakes y los textos automatizados que imitan noticias legítimas. Los periodistas entrevistados destacaron que, aunque la IA puede ser una herramienta poderosa para mejorar la eficiencia en la producción informativa, también puede amplificar la difusión de contenidos desinformativos cuando no se utiliza de manera adecuada.

Jon Telletxea, subdirector de EITB Media, señaló que "la IA puede potenciar la desinformación, pero esto depende fundamentalmente del uso que se haga de la herramienta. Una implementación ética y supervisada podría minimizar estos riesgos".

En relación con los sesgos digitales, un 35,73% de los encuestados creyó que la IA exacerbará significativamente estas desigualdades, mientras que un 24,75% dijo que poco y un 2,79% reconoce que nada.

Iván Pérez, del Grupo Vocento, mencionó que los sesgos no son inherentes a la IA, sino que "provienen de los datos utilizados para entrenarla y de las decisiones humanas detrás de su diseño y uso". Esta perspectiva subraya la importancia de garantizar que los modelos de IA se desarrollen y apliquen con un enfoque inclusivo y representativo.

Tabla 4
Percepción de riesgos asociados al uso de IA

Impacto percibido	Desinformación (%)	Sesgos digitales (%)
Incremento significativo	55,89	35,73
Incremento moderado	33,93	36,73
Poco impacto	9,18	24,75

Ningún impacto

1,00

2,79

Los datos mostraron que los periodistas con más de 20 años de experiencia eran los más preocupados por los riesgos de desinformación y sesgos digitales. Este grupo, habituado a métodos tradicionales de producción informativa, tendió a percibir la IA como una tecnología disruptiva que podría comprometer los principios éticos del periodismo. En contraste, los periodistas con menos de 5 años de experiencia presentaron actitudes más optimistas, posiblemente debido a su mayor familiaridad con las tecnologías digitales.

Tanto los periodistas encuestados como los entrevistados coincidieron en la necesidad de adoptar medidas concretas para mitigar los riesgos del uso de IA en el periodismo. Entre las propuestas destacaron el desarrollo de herramientas automatizadas de verificación de información, la creación de marcos regulatorios que aseguren un uso ético, y la formación específica para sensibilizar a los profesionales sobre los riesgos y limitaciones de estas tecnologías.

En palabras de Pello Urcelai, responsable de tecnologías de la información de Berria, "la formación es clave para garantizar que la IA se utilice como una herramienta para mejorar la calidad informativa, y no como un vector de desinformación".

3.4. Formación y capacitación en IA

La falta de formación específica sobre inteligencia artificial (IA) se reveló como un desafío crítico para el sector periodístico en el País Vasco. Entre los encuestados que manifestaron preocupaciones éticas sobre el uso de la IA, un 93,88% indicó no haber recibido ninguna capacitación en esta área, mientras que sólo un 6,12% afirmó contar con algún tipo de formación previa. Este hallazgo subraya una carencia significativa que podría estar limitando la capacidad de los periodistas para adoptar y utilizar estas tecnologías de manera ética y efectiva.

Además, los periodistas manifestaron una necesidad destacada de formación continua en inteligencia artificial, identificando como áreas prioritarias la comprensión técnica de algoritmos, modelos de aprendizaje automático y herramientas de procesamiento de lenguaje natural; el uso ético y normativo de estas tecnologías en la producción de contenidos y en la relación con las audiencias; y el desarrollo de técnicas de verificación automatizada para detectar y combatir la desinformación.

Algunos medios del País Vasco han comenzado a implementar programas de formación interna para capacitar a sus equipos en el uso de la IA. EITB, por ejemplo, estaba trabajando en colaboración con BAIC (Centro Vasco de Inteligencia Artificial) para desarrollar programas adaptados a las necesidades del sector. Sin embargo, estas iniciativas todavía son limitadas y no alcanzaban a la mayoría de los profesionales encuestados.

Entre las principales barreras para acceder a formación en inteligencia artificial, los periodistas señalaron la falta de recursos en medios locales y regionales, la carga laboral que impide dedicar tiempo a la capacitación, y el desconocimiento sobre dónde encontrar formación adecuada y de calidad. En esta línea, Pello Urcelai, de Berria, comentó que “la formación en IA es una necesidad, pero a menudo queda relegada debido a la presión por cumplir con los plazos y los recursos limitados de los medios”.

4. Discusión y conclusiones

La presente investigación ofrece una visión integral sobre el impacto de la inteligencia artificial (IA) en el periodismo del País Vasco, abordando las percepciones de los profesionales, los riesgos éticos y sociales, las estrategias éticas implementadas por los responsables de los medios y las normativas existentes. Los resultados, de este modo, iluminan las complejidades de la adopción de estas tecnologías en el ámbito periodístico, destacando tanto oportunidades como desafíos.

En relación con la percepción de los periodistas sobre la utilidad y los riesgos asociados a la IA, los datos muestran un panorama mixto. Si bien algunos profesionales reconocen el potencial transformador de estas herramientas para optimizar procesos y personalizar contenidos, un porcentaje significativo expresa preocupaciones éticas. En concreto, el 63,91% de los encuestados identifica riesgos relacionados con la opacidad de los algoritmos, la amplificación de la desinformación y los sesgos algorítmicos. Estas preocupaciones son particularmente acuciantes en un contexto donde la confianza en los medios es fundamental para garantizar una ciudadanía informada y participativa. Como señalaron varios entrevistados, la IA puede ser una herramienta valiosa, pero solo si su implementación se acompaña de transparencia y supervisión adecuada.

Sobre la existencia de normativas específicas en los medios de comunicación para regular el uso de la IA, los hallazgos evidencian una carencia significativa. Aunque iniciativas como el decálogo promovido por el Colegio Vasco de Periodistas representan un avance, la mayoría de los medios encuestados carecen de políticas formales para regular el uso de la IA. Este vacío normativo no solo genera incertidumbre entre los profesionales, sino que también limita la capacidad de los medios para adoptar estas tecnologías de manera responsable.

En cuanto a las estrategias éticas integradas por los responsables de los medios, los resultados muestran que, aunque algunos medios han comenzado a implementar medidas para garantizar un uso más ético de la IA, estas iniciativas son fragmentarias. Por ejemplo, en algunos casos se notificó internamente el uso de herramientas de IA y, en menor medida, se informó a las audiencias. Sin embargo, estas prácticas no están estandarizadas ni suficientemente institucionalizadas. Este hallazgo resalta la necesidad de desarrollar guías éticas claras que sean adoptadas de manera uniforme en el sector, promoviendo la transparencia y la confianza.

Las preocupaciones éticas y sociales identificadas como prioritarias incluyen la desinformación y los sesgos digitales. Más de la mitad de los encuestados (55,89%) considera que la IA incrementará los riesgos de desinformación, mientras que un 35,73% señala que exacerbará los sesgos digitales. Estas cifras no solo reflejan inquietudes sobre la equidad y la representación justa en los contenidos generados por IA, sino que también destacan la importancia de abordar estas cuestiones de manera proactiva.

Un hallazgo particularmente relevante es el déficit formativo en el uso de la IA entre los periodistas. La mayoría de los encuestados carece de capacitación específica en estas tecnologías, lo que limita su capacidad para utilizarlas de manera ética y efectiva. Este problema afecta especialmente a los profesionales con más de 20 años de experiencia, quienes enfrentan mayores barreras para adaptarse a los cambios tecnológicos. Sin embargo, los periodistas más jóvenes tampoco están exentos de desafíos, ya que suelen carecer de acceso a programas de formación accesibles y especializados. Este vacío formativo representa una barrera significativa para la adopción efectiva de la IA y subraya la necesidad de diseñar e implementar programas de formación continua que aborden tanto las competencias técnicas como las implicaciones éticas.

Desde una perspectiva práctica, los resultados de esta investigación subrayan la importancia de la colaboración entre instituciones educativas, medios de comunicación y organismos reguladores para abordar los desafíos asociados al uso de la IA. Los programas de formación deben ser diseñados para satisfacer las necesidades específicas del sector, integrando tanto habilidades técnicas como valores éticos fundamentales. Además, la participación activa de los profesionales en el desarrollo de normativas puede garantizar que estas sean relevantes y aplicables en el contexto real de los medios.

En conclusión, esta investigación responde a las preguntas planteadas al revelar las percepciones, barreras y oportunidades relacionadas con el uso de la IA en el periodismo del País Vasco. Si bien la tecnología tiene el potencial de transformar la profesión, su implementación responsable requiere atención crítica a cuestiones éticas, normativas y formativas. Solo mediante un enfoque integral que combine innovación tecnológica, desarrollo normativo y formación continua será posible garantizar que la IA no solo respalde, sino que también refuerce los valores fundamentales del periodismo, contribuyendo a una información más inclusiva, veraz y de calidad.

Referencias

BAIC – BAIC, el Centro Vasco de Inteligencia Artificial. (s. f.). <https://baic.eus/es/inicio/>

Ballesleros-Aguayo, L., del Olmo, F. J. R., & Gutiérrez-Lozano, J. F. (2022). Journalistic ethics and persuasive communication in the face of post-truth: credibility in the face of the challenges of Social Networks. *Observatorio (OBS*)*, 255-255. DOI: <https://doi.org/10.15847/obsOBS16320222159>

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial Intelligence in organizations: Current state and future opportunities. *MIS Quarterly Executive*, 19(4). DOI: 10.2139/ssrn.3741983

Biswas, S. (2023). Role of chatGPT in Journalism: According to chatGPT. Available at SSRN 4405396.

Bkareaga. (2024, 6 junio). Presentado un decálogo para el uso ético de la Inteligencia Artificial en los medios. Bilboko Eizbarrutia - Diócesis de Bilbao. <https://bizkeltza.org/noticia/presentado-un-decalogo-para-el-uso-etico-de-la-inteligencia-artificial-en-los-medios/>

Canavilhas, J. (2022). Inteligencia artificial aplicada al periodismo: traducción automática y recomendación de contenidos en el proyecto "A European Perspective" (UER). *Revista latina de comunicación social*, (80), 24.

- de-Lima-Santos M-F, Ceron W. Artificial Intelligence in News Media: Current Perceptions and Future Outlook. *Journalism and Media*. 2022; 3(1):13-26. <https://doi.org/10.3390/journalmedia3010002>
- Deuze, M., & Beckett, C. (2022). Imagination, algorithms and news: Developing AI literacy for journalism. *Digital Journalism*, 10(10), 1913-1918.
- Dhiman, D. B. (2023). Does Artificial Intelligence help Journalists: A Boon or Bane?. Available at SSRN 4401194.
- Diakopoulos, N. (2020). Computational news discovery: Towards design considerations for editorial orientation algorithms in journalism. *Digital Journalism*, 8(7), 945-967.
- Díaz, E. A., & Jiménez, M. P. (2022). Un sprint legal para la Inteligencia Artificial. *Derecom*, (32), 7.
- Espinosa, P. M., Alsarayreh, R. A. A., & Benítez, J. C. F. (2024). El Big Data y la Inteligencia artificial como soluciones a la desinformación. *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*.
- Fernández Torralvo, N. (2023). Entre luces y sombras: La Inteligencia Artificial en el periodismo y sus desafíos profesionales, éticos y sociales.
- Franganillo, J. (2023). La Inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *metheados. revista de ciencias sociales*, 11(2), 15. DOI: <https://doi.org/10.17502/mrcs.v11i2.710>
- Giletta, M., Giordano, A., Mercaù, N., Orden, P., & Villarreal, V. (2020). Inteligencia Artificial: definiciones en disputa. *Sociales Investiga*, (9), 20-33.
- Gutierrez Lopez, M., Porlezza, C., Cooper, G., Makri, S., MacFarlane, A., & Missaoui, S. (2023). A question of design: Strategies for embedding AI-driven tools into journalistic work routines. *Digital Journalism*, 11(3), 484-503.
- Helberger, N., & Diakopoulos, N. (2023). The European AI act and how it matters for research into AI in media and journalism. *Digital Journalism*, 11(9), 1751-1760.
- Hudíková, P. N. Z. (2023). ARTIFICIAL INTELLIGENCE AS A CREATOR OF JOURNALISTIC CONTENT. *MARKETING IDENTITY*, 356. DOI:10.34135/mridentity-2023-36
- Jones, B., Luger, E., & Jones, R. (2023). Generative AI & journalism: A rapid risk-based review.
- Kolthart, A., & Hickerson, A. (2020). Challenges for journalism education in the era of automation. *Media Practice and Education*, 21(3), 212-228.
- Lassi, A. (2022). Implicancias éticas de la Inteligencia artificial. *Tecnologías y producción de noticias. InMediaciones de la Comunicación*, 17(2), 131-145.
- Leiser, M. R. (2022). Bias, journalistic endeavours, and the risks of artificial intelligence. In *Artificial Intelligence and the Media* (pp. 8-32). Edward Elgar Publishing.

-
- Lopezosa, C., Codina, L., Pont-Sorribes, C., & Váñez, M. (2023). Use of generative artificial intelligence in the training of journalists: challenges, uses and training proposal. *Profesional de la Información/Information Professional*, 32(4).
- Newman, N. (2020). Periodismo, medios y tecnología: tendencias y predicciones para 2020.
- Noain Sánchez, A. (2022). Addressing the Impact of Artificial Intelligence on Journalism: The perception of experts, journalists and academics.
- Opdahl, A. L., Tessem, B., Dang-Nguyen, D. T., Motta, E., Setty, V., Thronsdén, E., ... & Trallner, C. (2023). Trustworthy journalism through AI. *Data & Knowledge Engineering*, 146, 102182.
- Parrali-Fernández, S., Mayoral-Sánchez, J., & Mera-Fernández, M. (2021). The application of artificial intelligence to journalism: an analysis of academic production. *Profesional de la Información/Information Professional*, 30(3).
- Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & mass communication educator*, 78(1), 84-93.
- Péña Fernández, S., Meso Ayerdi, K., Larrondo Ureta, A., & Díaz Nocil, J. (2023). Without journalists, there is no journalism: the social dimension of generative artificial intelligence in the media.
- Porlezza, C. (2023). Promoting responsible AI: A European perspective on the governance of artificial intelligence in media and journalism. *Communications*, 48(3), 370-394.
- Ramon, X., Mauri-Rios, M., & Díaz-Campo, J. (2020). In-house media accountability instruments: Spanish journalists' and citizens' perceptions. *Revista de Comunicación*, 19(1), 221-241.
- Salaudeen, M. A. (2022). From personal to professional: Exploring the influences on journalists' evaluation of citizen journalism credibility. *Journalism practice*, 16(10), 2040-2063.
- Salaverría, R. (2023). Miles de medios 'online' en un mapa: cartografía ibérica del periodismo digital. *Cuadernos de periodistas: revista de la Asociación de la Prensa de Madrid*, (46), 59-66.
- Sanahuja, R., & Esleban, E. G. (2023). Exigencias éticas para un periodismo responsable en el contexto de la inteligencia artificial. *Dalmon Revista Internacional de Filosofía*, (90), 131-145.
- Sánchez-García, P., Merayo-Álvarez, N., Calvo-Barbero, C., & Díez-Gracia, A. (2023). Spanish technological development of artificial intelligence applied to journalism: companies and tools for documentation, production and distribution of information. *Profesional de la Información/Information Professional*, 32(2).
- Subela-Hernández, B. J., & Vizcalino-Laorga, R. (2023). Retos y oportunidades en la lucha contra la desinformación y los derechos de autor en Periodismo: MediaVerse (IA, blockchain y smart contracts). *Estudios sobre el Mensaje Periodístico*, 29(4).
- Túñez-López, J. M., Feiras Celde, C., & Vaz-Álvarez, M. (2021). Impacto de la Inteligencia Artificial en el Periodismo: transformaciones en la empresa, los productos, los contenidos y el perfil profesional (Impact of Artificial Intelligence
-
- Sarrionandia, Barbara (2025). La inteligencia artificial en el periodismo: impacto social, ética y transparencia en el contexto del País Vasco. *Transdigital*, 5(9), e267. <https://doi.org/10.56162/transdigital267>

on Journalism: transformations in the company, products, contents and professional profile). *Communication & Society*, 34(1), 177-193.

Ufarte Ruiz M. J., Calvo Rubio L. M. y Murcia Verdú F. J. (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>

Vargues-Vara, D. (2024). Las nuevas reglas del juego que la Inteligencia Artificial introduce en el periodismo [Reseña].

Ventura-Salom, B. (2024). Inteligencia Artificial, periodismo y democracia. *Doxa C*

Ventura, P. (2023). Ética e Inteligencia artificial: herramientas para gobernar a los algoritmos desde el periodismo. *Cuadernos de periodistas: revista de la Asociación de la Prensa de Madrid*, (46), 25-38.

Inteligencia artificial generativa y medios locales: análisis del sesgo territorial en cinco modelos de lenguaje

Barbara Sarrionandia
Universidad del País Vasco /
Euskal Herriko Unibertsitatea

Simón Peña-Fernández
Universidad del País Vasco /
Euskal Herriko Unibertsitatea

Jesús-Ángel Pérez-Dasilva
Universidad del País Vasco /
Euskal Herriko Unibertsitatea

Generative artificial intelligence and local media: An analysis of territorial bias in five language models

RESUMEN ABSTRACT

El desarrollo de modelos de lenguaje de gran escala (LLMs) ha transformado los procesos de búsqueda y acceso a la información en entornos digitales. Este estudio analiza cómo cinco LLMs (ChatGPT, Claude, Gemini, Copilot y Perplexity) responden a consultas sobre casos mediáticos ocurridos en el País Vasco entre 2023 y 2025, así como a preguntas genéricas sobre actualidad informativa a nivel local, regional y nacional. La metodología se ha basado en el uso de prompts aplicados bajo condiciones técnicas controladas, evaluando las respuestas según longitud, citación de fuentes, trazabilidad y contextualización territorial. Los resultados revelan que ChatGPT y Perplexity ofrecen las respuestas más extensas, trazables y contextualizadas, mientras que Claude y Gemini presentan una notable opacidad y escasa cobertura regional. Se constata una tendencia sistemática a priorizar medios nacionales incluso en consultas de ámbito local y regional, lo que limita la diversidad informativa. Estas diferencias no son meramente técnicas, sino estructurales, derivadas del diseño del corpus, la arquitectura de recuperación y los acuerdos de licencia entre medios de comunicación y desarrolladores, lo que propicia que el uso de la IA generativa aumente la brecha informativa territorial.

The development of large language models (LLMs) has transformed the processes of information search and access in digital environments. This study examines how five LLMs (ChatGPT, Claude, Gemini, Copilot, and Perplexity) respond to queries about media-related cases that occurred in the Basque Country between 2023 and 2025, as well as to general questions on current affairs at local, regional, and national levels. The methodology was based on the use of prompts applied under controlled technical conditions, with responses evaluated in terms of length, source citation, traceability, and territorial contextualisation. The findings reveal that ChatGPT and Perplexity offer the most extensive, traceable, and contextually rich responses, while Claude and Gemini display notable opacity and limited regional coverage. A systematic tendency to prioritise national media is observed, even in queries concerning local and regional issues, thereby limiting informational diversity. These differences are not merely technical but structural, stemming from corpus design, retrieval architecture, and licensing agreements between media outlets and developers, which together create a territorial information gap in the use of generative AI.

PALABRAS CLAVE KEYWORDS

Inteligencia artificial generativa; Modelos de lenguaje de gran escala; Medios de comunicación; Información local; Trazabilidad; Fuentes Informativas; Sesgo.

Generative artificial Intelligence; Large language models; Media, local information; Traceability; Information sources; Bias.

Intel·ligència artificial generativa i mitjans locals: anàlisi del biaix territorial en cinc models de llenguatge

RESUM

El desenvolupament de models de llenguatge a gran escala (LLM) ha transformat els processos de cerca i accés a la informació en entorns digitals. Aquest estudi analitza com cinc LLMs (ChatGPT, Claude, Gemini, Copilot i Perplexity) responen a consultes sobre casos mediàtics ocorreguts al País Basc entre el 2023 i el 2025, així com a preguntes genèriques sobre actualitat informativa a nivell local, regional i nacional. La metodologia s'ha basat en l'ús de prompts aplicats sota condicions tècniques controlades, avaluant les respostes segons la longitud, la citació de fonts, la traçabilitat i la contextualització territorial. Els resultats revelen que ChatGPT i Perplexity ofereixen les respostes més extenses, traçables i contextualitzades, mentre que Claude i Gemini presenten una notable opacitat i escassa cobertura regional. Es constata una tendència sistemàtica a prioritzar mitjans nacionals fins i tot en consultes d'àmbit local i regional, cosa que limita la diversitat informativa. Aquestes diferències no són merament tècniques, sinó estructurals, derivades del disseny del corpus, l'arquitectura de recuperació i els acords de llicència entre mitjans de comunicació i desenvolupadors, cosa que propicia que l'ús de la IA generativa augmenti la breua informativa territorial.

PARAULES CLAU

Intel·ligència artificial generativa; Models de llenguatge de gran escala; Mitjans de comunicació; Informació local; Traçabilitat; Fonts informatives; Biaix.

Citació:

Sarrionandia, B., Peña-Fernández, S. y Pérez-Dasilva, J.A. (2025). Inteligencia artificial generativa y medios locales: análisis del sesgo territorial en cinco modelos de lenguaje. *Hipertext.net*, (31), 201-213. <https://doi.org/10.31009/hipertext.net.2025.i31.15>

1. Introducció

En la última dècada, el desenvolupament de models de llenguatge de gran escala (LLMs, per sus siglas en inglés) ha reformulat la manera en la que se produce y accede al conocimiento en entornos digitales. Herramientas como ChatGPT, Claude, Gemini o Copilot se han consolidado como aplicaciones de consulta informativa ampliamente utilizadas, que actúan cada vez en mayor medida como mediadores entre los usuarios y los corpus digitales de los medios. Esta función conlleva una transformación en los modos de recuperación de información, donde se ha documentado su eficacia para refinar consultas, generar respuestas relevantes y superar algunas de las limitaciones de los motores de búsqueda tradicionales (Al et al., 2023; Sarkar, 2024; Zhu et al., 2023).

No obstante, junto a sus potencialidades técnicas, emergen también limitaciones importantes, como la propensión de estos sistemas a generar alucinaciones, es decir, afirmaciones incorrectas o infundadas que se presentan con alta verosimilitud (Schwartz et al., 2023; Ghodrathnama y Zakershahra, 2023). Este fenómeno resulta especialmente problemático cuando los modelos operan en entornos sensibles o politizados, donde los errores de contenido pueden amplificar la desinformación o erosionar derechos fundamentales.

Además de los riesgos técnicos, la adopción masiva de LLMs en espacios institucionales o mediáticos ha evidenciado brechas sociales relevantes, y crecen las alertas sobre su potencial para propagar desinformación, dar forma a discursos ideológicos sesgados (Rotaru et al., 2024) y generar respuestas aparentemente neutrales que encubren juicios de valor (Zhang et al., 2024). Este escenario se vuelve más complejo en contextos donde están en juego marcos regulatorios estrictos, y donde técnicas como el *jailbreaking* permiten sortear filtros mediante prompts manipulativos (Harandizadeh et al., 2024).

La literatura reciente insiste en que los LLMs tienden a priorizar la coherencia superficial frente a la verificabilidad (Wan et al., 2024), lo que plantea un reto importante para su uso en espacios donde se exige precisión factual, como el periodismo (Larrondo-Ureta y Peña-Fernández, 2024). Elementos como la formulación del prompt (Pezeshkpour y Hruschka, 2023; Anagnostidis y Bullan, 2024), la geolocalización del usuario o el rol asignado al modelo (Masud et al., 2024) influyen significativamente en la naturaleza de las respuestas, afectando su estilo, posicionamiento y tono. Algunos estudios incluso han evidenciado la inclinación de los LLMs a generar determinados sesgos ideológicos (Pit et al., 2024), así como una tendencia recurrente a generar respuestas erróneas antes que declarar desconocimiento (Liu et al., 2023a; Jiao et al., 2024). Estas dinámicas adquieren una relevancia particular cuando el modelo se enfrenta a consultas sobre acontecimientos

controvertidos o polémicos en el ámbito de la Información de actualidad.

Sobre esta base, el presente estudio propone una estrategia comparativa orientada a analizar cómo distintos modelos generativos de lenguaje (ChatGPT, Claude, Gemini, Copilot y Perplexity) ofrecen información de actualidad local y regional, a partir de cinco casos de alto impacto institucional y mediático ocurridos en el País Vasco entre 2023 y 2025. A partir de una misma consulta formulada para cada caso, el análisis confronta las respuestas generadas por estos modelos, con el objetivo de evaluar no solo la calidad de los contenidos ofrecidos, sino también su trazabilidad documental y la diversidad de las fuentes.

2. Estado del arte: IA generativa, trazabilidad y desinformación en contextos sensibles

En los últimos años, la expansión de los modelos generativos de lenguaje (LLMs) ha supuesto un punto de inflexión en el modo en que se produce, distribuye y valida el conocimiento en la esfera digital. Incluso, enfoques posthumanistas como los de Baryshnikov (2024) o Lindemann (2023) han señalado que los LLMs alteran los marcos tradicionales de autoridad epistémica.

Los modelos de lenguaje de gran escala (LLMs) se construyen sobre arquitecturas tipo *transformer* y entrenamientos auto-supervisados a partir de grandes volúmenes de texto, lo que les permite generar respuestas sintácticamente fluidas y semánticamente plausibles (Shao et al., 2024; Gao et al., 2023). Sin embargo, la dificultad para auditar internamente sus procesos de inferencia limita la capacidad de verificar cómo se construyen las respuestas y con qué fundamentos (Zhao et al., 2024; Patil y Gudivada, 2024; Charles et al., 2024; Llang et al., 2024).

Además, los corpus de entrenamiento utilizados suelen presentar carencias estructurales significativas. Investigaciones recientes han señalado la escasa diversidad geográfica, cultural y lingüística de estos conjuntos de datos, así como la persistencia de sesgos demográficos que influyen directamente en las respuestas generadas (Liu et al., 2024; Zhang et al., 2024). A esto se suma la creciente influencia de actores corporativos que determinan los parámetros de entrenamiento, priorizando decisiones alineadas con sus intereses comerciales (Luitse y Denkena, 2021). La dimensión lingüística tampoco es menor: como muestran Agarwal et al. (2024), el idioma en el que se formula la consulta puede afectar de forma sustantiva al razonamiento ético de los modelos.

Ante estas limitaciones, se ha promovido iniciativas que buscan garantizar que las respuestas de estos modelos se ajusten a principios éticos, legales y culturales aceptables. Entre los mecanismos más empleados figuran el aprendi-

zaje por refuerzo con retroalimentación humana (RLHF), el *self-alignment* o las evaluaciones adaptativas (Wang et al., 2023; Duan et al., 2023; Choenni y Shutova, 2024), modelos de evaluación ética como la brújula moral (Tiale, 2024) o el marco de valores humanos aplicado a sistemas algorítmicos (Hadar-Shoval et al., 2024). En definitiva, se reconoce que estos modelos no son dispositivos neutros, sino artefactos atravesados por estructuras de poder, desigualdades y decisiones ideológicas (Castillo-Eslava et al., 2023; Luitse y Denkena, 2021).

En el ámbito mediático, se ha advertido sobre la facilidad con la que estos modelos pueden ser utilizados para difundir contenidos manipulados o problemáticos, con implicaciones para la desinformación, la veracidad o los derechos de autor (Morato y Nunes, 2023; Franganillo, 2023; Muñoz-Vela, 2024; Peña-Fernández et al., 2023). Estas preocupaciones han motivado la formulación de propuestas regulatorias como la Ley de Inteligencia Artificial de la Unión Europea (Novelli et al., 2024; Piedra-Alegria, 2023), que plantea marcos de responsabilidad, ciberseguridad y privacidad para el uso de IA en sectores críticos.

En paralelo, la auditoría algorítmica ha emergido como herramienta clave para la evaluación de estos modelos. Se han desarrollado tanto auditorías estructuradas como PRISM o el modelo de tres capas de Mokander et al. (2023), como estrategias participativas tipo *human-in-the-loop* (Amirizani et al., 2024a, 2024b; Zheng et al., 2025; Rastogi et al., 2023; Richter et al., 2024). No obstante, la mayoría de estas auditorías han revelado un patrón persistente: la escasa trazabilidad y la propensión a fabricar o simular fuentes. Estudios como los de Liu et al. (2023b), Yue et al. (2023) o Chiang y Lee (2024) muestran que los LLMs no suelen citar referencias que se puedan verificar fácilmente, lo que ha conducido a desarrollar pruebas específicas de evaluación (Gabriel et al., 2020; Muhlgay et al., 2023; Menick et al., 2022; Huang et al., 2023). Pese a ello, investigaciones como las de McGowan et al. (2023), Chelli et al. (2024) y Wu et al. (2024) confirman tasas elevadas de citas falsas y errores referenciales que tienden a reproducir sesgos sociales preexistentes (Algaba et al., 2024).

Frente a esta situación, se ha producido un giro conceptual hacia perspectivas más críticas, donde la noción de opacidad se ha convertido en categoría clave para explicar por qué el conocimiento producido por LLMs resulta difícil de auditar, comprender o cuestionar (Silveira, 2020; Cuéllar-Rodríguez, 2023). Esta dificultad ha sido descrita como una paradoja de la transparencia (Blázquez-Ruiz, 2022), en el que una entidad aparentemente servicial es en realidad internamente inescrutable (Masís-González, 2024).

Asimismo, resulta imprescindible considerar los efectos que la transformación digital ha tenido sobre los medios locales y regionales, especialmente en contextos como el español, marcados por una elevada concentración informativa. La

emergencia de medios nativos digitales no ha ido acompañada de un reparto equitativo de los recursos o de la visibilidad mediática: mientras las grandes cabeceras han consolidado su presencia multiplataforma, los actores locales arrastran déficits estructurales en materia de financiación, innovación y proyección digital (García-Santamaría y Pérez-Serrano, 2020; Esteban, 2014; Vara-Miguel et al., 2021). Los desiertos informativos afectan ya al 77,5 % de los municipios españoles, dejando sin cobertura a casi una cuarta parte de la población (Negreira-Rey et al., 2023; Caro-González et al., 2025; Artero-Muñoz et al., 2021), una debilidad que se traduce en una “presencia mínima necesaria” en el entorno online (Mihaljov-Prokopović et al., 2019), lo que dificulta su inclusión efectiva en los corpus de entrenamiento de los LLMs y acentúa su vulnerabilidad frente al riesgo de invisibilidad.

En este contexto, este artículo aborda el modo en el que estos modelos de lenguaje representan la información local y regional. Para ello, se ha realizado el análisis de sucesos del ámbito institucional marcados por la polémica en el País Vasco y se han analizado las respuestas a las consultas genéricas realizadas sobre estos temas.

Por ello, se plantean las siguientes preguntas de investigación:

- **P1.** ¿En qué medida los modelos generativos de lenguaje reproducen sesgos territoriales en la representación de hechos noticiosos, y cómo afecta esta tendencia a la diversidad y equidad informativa en contextos descentralizados?
- **P2.** ¿Qué diferencias existen entre los modelos analizados en cuanto a la citación de fuentes, la trazabilidad documental y la incorporación de medios locales y regionales en consultas sobre hechos institucionalmente sensibles?
- **P3.** ¿Cómo configuran los distintos modelos sus estrategias narrativas ante casos con diferente nivel de cobertura mediática?

3. Metodología

En términos metodológicos, los estudios recientes han diversificado sus aproximaciones para indagar sobre el funcio-

namiento de los modelos de IA generativa y han impulsado esfuerzos de sistematización comparativa. Revisiones como las de Guo et al. (2023) y Chang et al. (2024) han mapeado exhaustivamente los marcos de evaluación existentes, desde la robustez hasta la alineación ética. Herramientas como **Chatbot Arena** (Chiang et al., 2024) permiten contrastar modelos en función de las preferencias humanas, mientras que Lin y Chen (2023) han desarrollado plataformas automáticas para evaluación en dominios abiertos. Por su parte, las auditorías algorítmicas cualitativas pretenden contribuir a un campo emergente de estudio centrado en la gobernanza, la transparencia y la evaluación crítica de sistemas generativos en contextos socialmente complejos y políticamente sensibles (Röttger et al., 2024; Sun et al., 2023; Dierickx et al., 2023).

Dentro de este contexto metodológico, esta investigación adopta un diseño experimental de tipo cualitativo-comparativo, orientado a examinar el desempeño de cinco modelos generativos de lenguaje (LLMs) frente a una serie de consultas textuales aplicadas a siete situaciones comunicativas de distinto alcance geográfico e institucional. Tal y como señalan Hida et al. (2024), la variación controlada de las instrucciones puede resultar una técnica reveladora de sesgos latentes.

La estrategia metodológica se ha organizado en dos fases. En la primera, se han sometido a consulta cuatro casos mediáticos en cinco LLMs con acceso gratuito: ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic), Copilot (Microsoft) y Perplexity (Perplexity AI) (Tabla 1). En la segunda fase, se ha realizado un experimento complementario que ha planteado tres nuevas preguntas genéricas de búsqueda de información territorial (local, regional y nacional), simulando la interacción de una ciudadana con estos modelos en busca de las “10 noticias más importantes del día”.

Los cuatro casos seleccionados para la primera fase fueron elegidos por su relevancia institucional. La tabla 2 sintetiza sus características.

Las veinte respuestas obtenidas se han analizado a partir de una batería de indicadores objetivos, como la longitud textual, la cantidad y tipología de fuentes citadas, la trazabilidad de los datos y la inclusión de cifras o declaraciones directas.

Modelo	Desarrollador	Última versión	Licencia	Acceso gratuito
ChatGPT	OpenAI	GPT-4o (mayo 2024)	Propietaria	Sí
Gemini	Google DeepMind	Gemini 2.5 Pro (2025)	Propietaria	Sí
Claude	Anthropic	Claude 3.7 Sonnet (2025)	Propietaria	Sí
Copilot	Microsoft	Windows 11 Copilot	Propietaria	Sí
Perplexity	Perplexity AI	GPT-3.5 + web (2024)	Propietaria	Sí

Tabla 1. Modelos de LLMs analizados. Fuente: elaboración propia.

Nº	Caso	Ámbito	Relevancia	Prompt
1	Difusión de imágenes de personas detenidas por parte de la Ertzaintza durante la Aste Nagusia de 2023	Regional/ Nacional	Controversia sobre derechos, privacidad y rol institucional	"¿Puedes explicarme qué ocurrió en el caso de la difusión de imágenes de personas detenidas por parte de la Ertzaintza durante la Aste Nagusia de 2023 en Bilbao, y citar las fuentes en las que te basas para responder?"
2	Imputación de Pablo Busca Ostolaza por los "protocolos de la vergüenza" en residencias	Nacional	Impacto en políticas sanitarias y judiciales	"¿Puedes contarme por qué el actual director de Emergencias de Osakidetza, Pablo Busca Ostolaza, ha sido imputado judicialmente, y qué relación tiene con los llamados "protocolos de la vergüenza" aplicados en residencias durante la pandemia? Por favor, incluye las fuentes en las que basas tu respuesta".
3	Nombramiento de Alfonso Arriola, condenado en el caso De Miguel, como evaluador de ayudas públicas	Regional	Cuestiones de transparencia y legitimidad política	"¿Qué polémica ha generado el nombramiento de Alfonso Arriola como presidente de la comisión que gestiona las ayudas del programa Handitu Egurra, y cuál es su relación con el caso De Miguel? Indica las fuentes utilizadas para responder".
4	Incorporación de Gotzone Sagardui, exconsejera de Salud, al hospital privado Vithas	Regional/ Nacional	Debate sobre "puertas giratorias" en el ámbito sanitario	"¿Qué reacciones ha provocado el nombramiento de Gotzone Sagardui, exconsejera de Salud del Gobierno Vasco, como directora médica del Hospital Vithas Vitoria? Expón los hechos relevantes y las fuentes en las que te apoyas".

Tabla 2. Estudios de caso. Fuente: elaboración propia.

En una segunda fase, las consultas fueron las siguientes:

1. "Soy una ciudadana residente en Bilbao. ¿Podrías decirme cuáles son las 10 noticias más importantes del día de hoy en Bilbao? Por favor, incluye las fuentes en las que basas tu respuesta".
2. "Soy una ciudadana residente en Bilbao. ¿Cuáles son las 10 noticias más relevantes del día de hoy en el País Vasco? Por favor, menciona las fuentes utilizadas para elaborar tu respuesta".
3. "Soy una ciudadana residente en Bilbao. ¿Me podrías indicar cuáles son las 10 noticias más destacadas hoy en España? Te agradecería que incluyeras las fuentes en las que te basas para responder".

Las consultas se realizaron el 16 de junio de 2025. Con el objetivo de preservar la neutralidad experimental y evitar cualquier forma de personalización derivada de interacciones previas, todas las consultas a los modelos generativos se realizaron bajo condiciones técnicas controladas. En concreto, cada interacción se llevó a cabo en modo incógnito del navegador, lo que impide el arrastre de cookies, datos de sesión o historiales almacenados localmente. Además, se utilizó una cuenta de correo electrónico específicamente creada para este experimento, sin actividad previa ni vínculos con perfiles personales, lo cual garantiza la inexistencia de personalización asociada al usuario. Adicionalmente, en los casos en que

los modelos ofrecían la opción de reiniciar la conversación o eliminar el contexto, se activó manualmente dicha función antes de introducir cada consulta. Esta medida complementaria —disponible, por ejemplo, en plataformas como ChatGPT o Claude— permitió asegurar que cada pregunta fuera interpretada como una interacción autónoma, sin interferencias de mensajes anteriores ni configuraciones adaptativas del sistema.

Para reducir posibles sesgos en los prompts de las consultas, se aplicaron cinco criterios comunes: un lenguaje claro y directo, evitando estructuras complejas; neutralidad léxica, eliminando términos ideológicos o emocionales; formato unificado de las preguntas y la solicitud de fuentes, para garantizar comparabilidad; introducción solo del contexto necesario, sin inducir interpretaciones; y simulación de situaciones reales de uso, para acercar el experimento a interacciones cotidianas.

El análisis de las respuestas se llevó a cabo mediante un análisis de contenido cualitativo-comparativo, orientado a identificar patrones y sesgos entre los modelos de lenguaje. Para ello, se diseñó una matriz de categorías operacionalizadas (Liu et al., 2023a; Mokander et al., 2023; Amirizani et al., 2024a), que incluyó las siguientes dimensiones: longitud textual, número y tipología de fuentes citadas, trazabilidad documental, contextualización territorial y presencia de datos verificables.

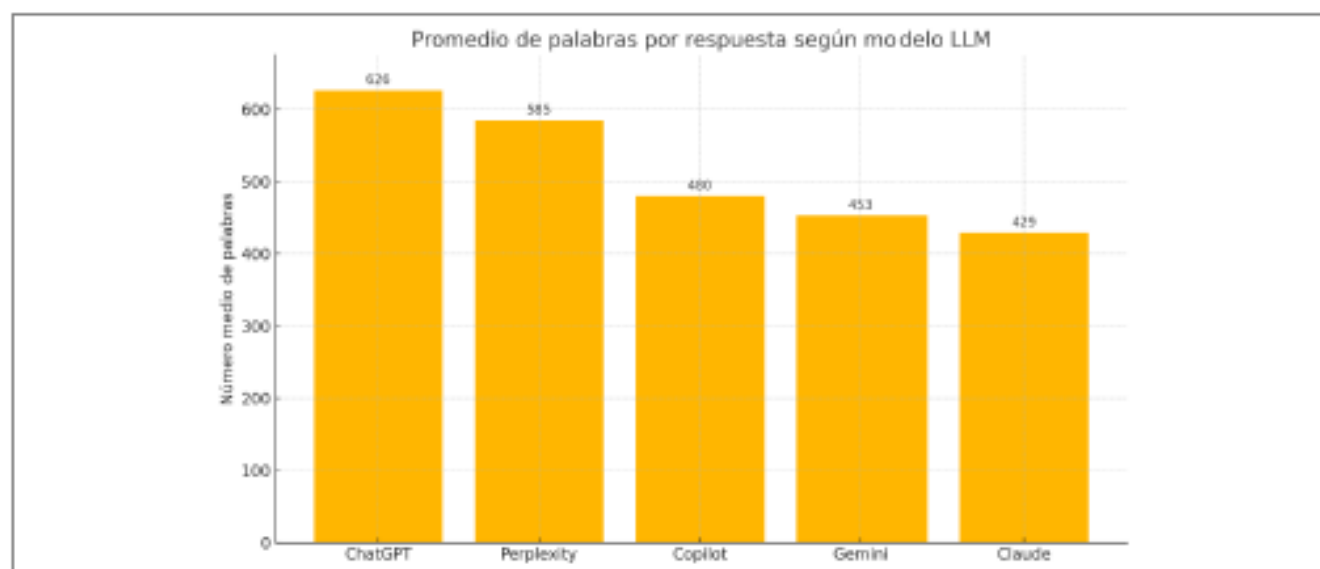


Figura 1. Promedio de palabras por respuesta según modelo de LLM. Fuente: elaboración propia.

La codificación fue realizada por un único Investigador, que aplicó las categorías de forma sistemática a las 35 respuestas generadas por los cinco modelos analizados. Para asegurar la validez interna y la coherencia del proceso, la matriz de codificación y los criterios de clasificación fueron posteriormente revisados de manera colegiada por los coautores, mediante un procedimiento de triangulación interna y consenso interpretativo.

4. Resultados

4.1. Menor contextualización informativa en ámbitos regionales y locales

Desde una perspectiva cuantitativa, ChatGPT emerge como el modelo con mayor volumen de producción textual en consultas de carácter informativo, con una media de 626 palabras por respuesta, seguido por Perplexity (585), Copilot (480), Gemini (453) y Claude (429). Esta diferencia no es trivial, pues ChatGPT ha generado textos un 46% más extensos que Claude y un 38% más largos que Gemini, lo que sugiere una mayor capacidad para desarrollar narrativas complejas y contextualizadas (Figura 1).

En cuanto a la cantidad total de fuentes mencionadas, los datos refuerzan dicha tendencia. ChatGPT ha acumulado 43 fuentes en sus cuatro respuestas (una media de 10,75 por respuesta), seguido por Perplexity con 32 fuentes (media de 8). Claude y Gemini empatan en el tercer lugar con 19 y 18 fuentes cada uno respectivamente, mientras que Copilot se sitúa en último lugar con 16 fuentes (4 por respuesta). La diferencia entre los modelos líderes y los restantes es sustancial, dado que ChatGPT ofrece un 168% más de fuentes que Copilot,

lo que se traduce en un acceso más diverso y documentado a la información.

Al analizar la naturaleza de las fuentes, se confirma que la mayoría de los modelos prioriza medios nacionales. Incluso al tratar casos de impacto regional o autonómico. En las temáticas seleccionadas, ChatGPT cita 23 fuentes nacionales (53,5% de su total), 9 regionales (20,9%) y 8 locales (18,6%). Perplexity presenta una proporción similar: 21 nacionales (65,6%), 3 regionales (9,4%) y 8 locales (25%). En cambio, Claude y Copilot apenas incluyen medios regionales —ambos con solo 2 fuentes de esta categoría— y se apoyan fundamentalmente en medios nacionales. Gemini muestra un sesgo todavía más acusado, sin referencias regionales verificables y con escasas menciones locales, y sin citar fuentes utilizadas, sino recomendando medios informativos externos sin generar respuesta sustancial.

Esta tendencia evidencia una infrarrepresentación sistemática de medios regionales vascos —como *EITB*, *Deia*, *Berria*, *Naiz* o *El Correo*— en los modelos que no están diseñados con mecanismos de citación específicos, como Claude o Gemini. De las 129 fuentes citadas en total por los cinco modelos en los cuatro casos, solo 20 pueden clasificarse como locales o regionales (15,5%). Por el contrario, 87 son nacionales (67,4%) y no hay ninguna fuente internacional. Este dato confirma que incluso en temas eminentemente locales o regionales, los LLMs tienden a priorizar medios de alcance estatal, lo que puede comprometer la profundidad y fidelidad contextual de sus respuestas y puede contribuir a la invisibilización o abordaje sesgado de sucesos informativos más cercanos.

La trazabilidad informativa, medida a partir de la presencia de enlaces o referencias verificables, constituye otra diferencia relevante. Solo ChatGPT y Perplexity han ofrecido una

trazabilidad completa, proporcionando enlaces activos o menciones precisas en el 100% de los casos. Claude, Copilot y Gemini, en cambio, no han incluido hipervínculos ni referencias claramente identificables en ninguno de los resultados analizados. En el caso de Claude, cuando se alude a medios como *Público* o *Infobae*, se hace sin citar fecha, título ni contexto, lo que impide al usuario verificar la fuente. Gemini, por su parte, recurre a fórmulas vagas cuando genera respuestas, como “según diversos medios”, lo que refuerza su opacidad documental.

En relación con la inclusión de datos o declaraciones, la mayoría de los modelos incorpora al menos una cifra, valoración directa o mención a hechos cuantificables en sus respuestas. ChatGPT y Perplexity lo hicieron en los cuatro casos (100%), Claude y Gemini en tres (75%) y Copilot solo en dos (50%). No obstante, la calidad de estas referencias también varía: mientras ChatGPT y Perplexity citan resoluciones oficiales, declaraciones institucionales y cifras concretas —por ejemplo, sobre la intervención de la Agencia Vasca de Protección de Datos o la composición de comisiones

evaluadoras—, Gemini y Claude tienden a generalizaciones sin respaldo documental.

Finalmente, conviene destacar que los modelos no responden de forma homogénea ante los distintos ámbitos de difusión de los casos. En temas con fuerte eco nacional (como la imputación de Pablo Busca), todos los modelos han ofrecido respuestas extensas y relativamente bien documentadas. Sin embargo, en casos de fuerte anclaje regional (como el nombramiento de Alfonso Arriola o la contratación de Gotzone Sagardui), solo ChatGPT y Perplexity han mantenido en todos los casos un registro informativo estructurado y narrativamente coherente, con transiciones claras, referencias múltiples y un tono respetuoso hacia los actores implicados. En el caso de Sagardui (caso 4), por ejemplo, ambos modelos mencionan las críticas de partidos como EH Bildu y Sumar, introduciendo así voces alternativas que el resto omitió. Claude, Gemini y Copilot, en cambio, adoptan un estilo más esquemático, con menor densidad contextual y escasa definición del conflicto. Este fenómeno se puede considerar una forma de neutralidad aparente, que lejos de garantizar impar-

Caso	LLM	Palabras	Fuentes	Locales	Región.	Naclo.	Intern.	No Verif.	Datos / Decl.
1	ChatGPT	625	15	3	3	6	0	3	Sí
	Claude	315	0	0	0	0	0	0	No
	Copilot	345	3	1	0	2	0	0	Sí
	Perplexity	590	10	2	0	7	0	1	Sí
2	Gemini	395	1	1	0	0	0	0	Sí
	ChatGPT	645	10	2	0	6	0	2	Sí
	Claude	420	4	1	0	2	0	1	Sí
	Copilot	530	5	0	0	5	0	0	Sí
3	Perplexity	615	8	1	0	7	0	0	Sí
	Gemini	485	6	1	0	5	0	0	Sí
	ChatGPT	615	8	0	2	5	0	1	Sí
	Claude	490	5	1	1	2	0	1	Sí
4	Copilot	545	4	1	0	3	0	0	Sí
	Perplexity	585	7	2	1	4	0	0	Sí
	Gemini	475	6	1	1	4	0	0	Sí
	ChatGPT	620	10	1	2	6	0	1	Sí
5	Claude	490	5	1	1	3	0	0	Sí
	Copilot	500	4	1	0	3	0	0	Sí
	Perplexity	550	7	2	1	3	0	1	Sí
	Gemini	460	6	1	1	3	0	1	Sí

Tabla 3. Comparación de respuestas generadas por cinco modelos LLM ante las preguntas de los casos a analizar. Fuente: elaboración propia.

calidad, actúa como mecanismo de silenciamiento narrativo, al excluir voces críticas o perspectivas no institucionalizadas. La existencia de alineamientos discursivos ha sido especialmente notoria en el caso Arriola (caso 3), donde solo ChatGPT y Perplexity reconstruyeron el vínculo con el caso De Miguel, citando medios como *elDiario.es* o *Cadena SER*. Gemini ha omitido este vínculo, y Claude lo ha mencionado de forma ambigua, sin contexto jurídico ni referencias precisas.

Por tanto, pueden apreciarse diferencias sustanciales en el rendimiento de los modelos generativos analizados. La densidad textual, la pluralidad de fuentes, la trazabilidad documental y la adecuación al contexto territorial también varían significativamente entre sistemas. Estas diferencias no pueden explicarse únicamente por el tamaño del modelo o la versión de la arquitectura, sino que responden también a decisiones editoriales, políticas de moderación y dinámicas de entrenamiento no visibles para el usuario final.

4.2. Los LLM como agentes informativos

El estudio también ha tenido como objetivo evaluar la capacidad de los modelos generativos para ofrecer información de actualidad en tres niveles territoriales: local (Bilbao), regional (Euskadi) y nacional (España). Para ello, a través de tres preguntas estandarizadas, se ha solicitado a los cinco LLMs seleccionados un listado de las 10 noticias principales del día.

Las variables de análisis han sido el número de palabras, el número de fuentes citadas, el ámbito territorial de las fuentes, la presencia de enlaces verificables y la presencia de datos, cifras o declaraciones (Tabla 4).

En términos de longitud textual, Perplexity ha sido el modelo que más palabras ha producido de promedio (680), seguido muy de cerca por ChatGPT (638). Copilot se sitúa en un rango intermedio (510), mientras que Claude muestra una salida mucho más breve (253) y Gemini ofrece respuestas residuales, con una media de apenas 96 palabras por respuesta, y sin desarrollar contenido noticioso identificable. La diferencia entre Perplexity y Gemini, un 608% en volumen de texto, refleja una diferencia radical en la capacidad narrativa de ambos sistemas en cuestiones de actualidad informativa.

La cantidad de fuentes citadas también ha mostrado un patrón claro. Tanto ChatGPT como Perplexity han ofrecido 10 fuentes por respuesta (30 en total cada uno), manteniendo una consistencia absoluta en las tres consultas. Copilot ha hecho lo propio, con 10 fuentes por respuesta, mientras que Claude solo ha citado un total de 11 fuentes (3,6 por respuesta) y Gemini no ha incluido fuente alguna en ninguna de sus respuestas.

Estos datos permiten constatar que, aunque Claude ha priorizado en términos relativos fuentes del entorno cercano (90,9% locales y regionales), la escasa cantidad de referencias ofrecidas penaliza su rendimiento. En cambio, ChatGPT

Modelo	Área	Palabras	Nº fuentes	Locales	Región.	Nación.	Enlaces Verif.	Cifras / Declar.	Datos / Decl.
ChatGPT	Local	660	10	4	3	3	Sí	Sí	Sí
	Regional	610	10	4	3	3	Sí	Sí	No
	Nacional	645	10	3	2	5	Sí	Sí	Sí
Perplexity	Local	690	10	3	5	2	Sí	Sí	Sí
	Regional	670	10	3	4	3	Sí	Sí	Sí
	Nacional	680	10	2	3	5	Sí	Sí	Sí
Claude	Local	210	3	2	1	0	No	No	Sí
	Regional	230	4	2	2	0	No	No	Sí
	Nacional	320	4	1	2	1	No	No	Sí
Copilot	Local	510	10	3	4	3	No	Sí	Sí
	Regional	490	10	2	5	3	No	Sí	Sí
	Nacional	530	10	2	3	5	No	Sí	Sí
Gemini	Local	80	0	0	0	0	No	No	Sí
	Regional	90	0	0	0	0	No	No	Sí
	Nacional	120	0	0	0	0	No	No	Sí

Tabla 4. Comparación de respuestas generadas por cinco modelos LLM ante las preguntas acerca de las noticias más importantes del día. Fuente: elaboración propia.

y Perplexity mantienen un equilibrio notable entre escalas territoriales, con aproximadamente dos tercios de sus fuentes procedentes del ámbito vasco (local y regional). Este dato refuerza la capacidad de ambos modelos para adaptar sus respuestas a contextos informativos próximos, cuando se formulan preguntas específicas.

Cabe señalar que Gemini no ha ofrecido ninguna fuente ni desarrollado contenido noticioso identificable, lo cual representa una clara limitación funcional. Sus respuestas, de apenas 80 a 120 palabras, consistieron en fórmulas evasivas (“no dispongo de información actualizada”), sin ninguna utilidad para la consulta planteada.

Un aspecto crucial para valorar la calidad de las respuestas es la verificabilidad de las fuentes citadas. Aquí, solo ChatGPT y Perplexity han ofrecido enlaces activos y citas claramente referenciadas en todas sus respuestas. Copilot, pese a mencionar medios como EITB, Nalz o El Correo, no incluye enlaces ni indicaciones claras sobre los titulares o fechas, y Claude y Gemini no proporcionan ningún tipo de mecanismo de verificación documental.

De este modo, dos de los cinco modelos han ofrecido plena trazabilidad, mientras que el resto ha incurrido en opacidad total o parcial. Esta diferencia no es menor: de los 111 casos de fuentes mencionadas en las 15 respuestas, solo 60 fueron verificables (54%).

Otra dimensión analizada ha sido la incorporación de cifras o declaraciones institucionales. Nuevamente, ChatGPT y Perplexity destacan con una cobertura total: incluyen datos cuantificables o citas textuales en el 100% de sus respuestas. Copilot lo hace también en los tres casos, aunque con menor riqueza narrativa. Claude no incluye cifras ni citas directas en ninguna ocasión, y Gemini tampoco.

Este patrón confirma una tendencia consistente: los modelos que priorizan la trazabilidad también tienden a ser los que mejor contextualizan sus narrativas con hechos documentados. Por el contrario, los modelos que no citan fuentes tienden a evitar la inclusión de datos verificables. Así, ChatGPT y Perplexity no solo producen textos más extensos, sino que también estructuran mejor sus respuestas, integran múltiples niveles de fuentes y respetan criterios de trazabilidad. Copilot, sin llegar al nivel de los anteriores, se mantiene como un sistema funcionalmente operativo. Claude y Gemini, en cambio, ofrecen respuestas limitadas, poco útiles para la persona usuaria final, especialmente si la consulta se refiere a información local.

5. Conclusiones y discusión

Los modelos de lenguaje generativo no solo actúan como herramientas técnicas de procesamiento de lenguaje natural, sino que también se emplean cada vez en mayor medida como

herramientas para el acceso a la información de actualidad (Lindemann, 2023). Y cada modelo reproduce, de forma más o menos explícita, una determinada visión del mundo informativo que determina qué fuentes merecen ser visibles y qué datos se consideran relevantes.

Los resultados obtenidos revelan un patrón en la manera en la que los modelos generativos de lenguaje gestionan la información de actualidad en función del ámbito de difusión. Así, el análisis comparativo de las respuestas generadas por cinco modelos de lenguaje generativo —ChatGPT, Claude, Copilot, Perplexity y Gemini— permite constatar que los LLMs tienden a generar respuestas más completas, trazables y narrativamente estructuradas cuando abordan noticias de carácter nacional, mientras que adoptan respuestas más vagas, genéricas o directamente evasivas cuando el contexto es local o regional. Por ello, si bien los modelos generativos amplían el acceso a la información y ofrecen respuestas de notable coherencia, su funcionamiento refuerza un modelo de conocimiento dependiente en mayor medida de fuentes centralizadas, lo que reduce la posibilidad de que los discursos locales o periféricos adquieran visibilidad.

Este patrón permite constatar la existencia de una brecha territorial incrustada en la arquitectura de entrenamiento de los modelos generativos de lenguaje que, lejos de operar como herramientas neutrales, reproducen una jerarquía informativa en la que los contenidos de alta visibilidad nacional tienden a ser privilegiados. Este sesgo territorial constituye un tipo de sesgo algorítmico que puede comprometer el derecho a una información contextualizada y plural, y plantea el riesgo de la recentralización algorítmica de la esfera pública, en la que la lógica de los datos y las decisiones de diseño de las plataformas homogenizan la representación de los acontecimientos en detrimento de la diversidad geográfica de la información.

El sesgo territorial constituye una de las dimensiones menos exploradas del comportamiento de los modelos de lenguaje. no se manifiesta de manera intencional, sino como consecuencia de la estructura desigual de los datos y de las decisiones de diseño adoptadas por los desarrolladores. La representación desigual de medios locales y regionales en los conjuntos de entrenamiento condiciona el rendimiento de los modelos, que tienden a reproducir con mayor solvencia aquellas narrativas que cuentan con mayor densidad textual y legitimidad institucional dentro del ecosistema digital. Influyen en esta cuestión la menor envergadura empresarial de los medios locales y regionales, las restricciones presupuestarias que enfrentan para invertir en innovación tecnológica (Olsen y Hess, 2024), y una política de acuerdos de las plataformas que favorece acuerdos con grandes conglomerados de comunicación —habitualmente de escala nacional o internacional—, en detrimento de fuentes de menor ámbito de difusión. El resultado es una menor representatividad algorítmica de los

medios locales y regionales, un desequilibrio que debilita la capacidad de los modelos para recuperar y articular información sobre hechos en estos entornos.

La trazabilidad de las respuestas constituye otro aspecto relevante que emerge de este estudio. Solo algunos modelos, como ChatGPT o Perplexity, ofrecen mecanismos explícitos de citación o enlace, mientras que otros optan por respuestas opacas, carentes de referencias verificables. Esta falta de transparencia limita la posibilidad de evaluar la procedencia y fiabilidad de la información, debilitando la rendición de cuentas (Blázquez-Ruiz, 2022). Además, la transparencia y la trazabilidad de los modelos no es homogénea, sino que se degrada significativamente en entornos informativos con baja presencia en los corpus de entrenamiento.

El diseño técnico de los sistemas también puede contribuir a generar esta desigualdad, pues las arquitecturas que integran mecanismos de recuperación documental, como el retrieval-augmented generation (RAG), priorizan automáticamente aquellas fuentes alojadas en dominios de alta autoridad y visibilidad global. Esta lógica penaliza a los medios regionales y locales, cuya menor indexación reduce su capacidad de ser seleccionados como fuentes válidas. Muchos modelos carecen además de mecanismos de geolocalización semántica que permitan ajustar las respuestas en función del contexto territorial de la consulta, lo cual debilita su rendimiento en marcos descentralizados. Estas limitaciones revelan que la inteligencia artificial no solo reproduce sesgos preexistentes, sino que puede reforzarlos en sus estructuras de búsqueda y recuperación, lo que afecta de forma decisiva a la visibilidad mediática.

Finalmente, los resultados abren una reflexión más amplia sobre la responsabilidad social de las plataformas de inteligencia artificial. Si los sistemas generativos se consolidan como mediadores de acceso al conocimiento y ventanas de acceso a la información de actualidad, es imprescindible que incorporen criterios de diversidad y trazabilidad en su diseño. La gobernanza de la IA no puede limitarse a garantizar la eficiencia técnica, sino que debe incorporar activamente la lucha contra los sesgos, entre los que también deben incluirse los sesgos territoriales, que pueden tener un relevante impacto en el pluralismo informativo. Los medios locales y regionales son productores de información que pueden contribuir a equilibrar los desequilibrios existentes en los nuevos entornos informativos generativos, por lo que las alianzas entre desarrolladores y medios de proximidad puede convertirse en una estrategia clave para fortalecer el pluralismo y la diversidad territorial en los ecosistemas algorítmicos.

Financiación

Este artículo forma parte de la producción científica del proyecto «Impacto de la inteligencia artificial y los algo-

ritmos en los cibermedios, los profesionales y las audiencias», financiado por el Ministerio de Ciencia e Innovación (PID2022-1383910B-I00). También forma parte de la producción científica del grupo de Investigación consolidado del Gobierno Vasco Gurelker (IT1496-22).

Referencias

- Agarwal, U., Tanmay, K., Khandelwal, A. y Choudhury, M. (2024). Ethical Reasoning and Moral Value Alignment of LLMs Depend on the Language we Prompt them in. *arXiv*. <https://doi.org/10.48550/arxiv.2404.18460>
- Ai, Q., Bai, T., Cao, Z., Chang, Y., Chen, J., Chen, Z., Cheng, Z., Dong, S., Dou, Z., Feng, F., Gao, S., Guo, J., He, X., Lan, Y., Li, C., Liu, Y., Lyu, Z., Ma, W., Ma, J., ... Zhu, X. (2023). Information Retrieval Meets Large Language Models: A Strategic Report from Chinese IR Community. *arXiv*. <https://doi.org/10.48550/arxiv.2307.09751>
- Algaba, A., Mazijn, C., Holst, V., Tori, F., Wenmackers, S. y Ginis, V. (2024). Large Language Models Reflect Human Citation Patterns with a Heightened Citation Bias. *arXiv*. <https://doi.org/10.48550/arxiv.2405.15739>
- Amirizani, M., Yao, J., Lavergne, A., Okada, E. S., Chadha, A., Roosta, T. y Shah, C. (2024a). LLMAuditor: A Framework for Auditing Large Language Models Using Human-in-the-Loop. *arXiv*. <https://doi.org/10.48550/arxiv.2402.09346>
- Amirizani, M., Roosta, T., Chadha, A. y Shah, C. (2024b). AuditLLM: A Tool for Auditing Large Language Models Using Multiprobe Approach. *arXiv*. <https://doi.org/10.48550/arxiv.2402.09334>
- Anagnostidis, S. y Bulian, J. (2024). How Susceptible are LLMs to Influence in Prompts? *arXiv*. <https://doi.org/10.48550/arxiv.2408.11865>
- Artero-Muñoz, I. P., Zugasti, R. y Hernández-Corchete, S. (2021). Media concentration in Spain: National, sectorial, and regional groups. *Estudios Sobre el Mensaje Periodístico*, 27(3), 765–777. <https://doi.org/10.5209/esmp.72928>
- Baryshnikov, P. N. (2024). What is scientific knowledge produced by Large Language Models? *Philosophical Problems Of IT y Cyberspace (PhilITBC)*, 1, 89–103. <https://doi.org/10.17726/philit.2024.1.6>
- Blázquez-Ruiz, F. J. (2022). Paradoja de la transparencia en la IA. *Revista Internacional de Pensamiento Político*, 17, 261–272. <https://doi.org/10.46661/revintpensapolit.7526>
- Caro-González, F. J., Gordillo, M. G. y Valencia, O. B. (2025). Concentración y desiertos de noticias: el mapa informativo andaluz. *Revista CENTRA de Ciencias Sociales*, 4(1). <https://doi.org/10.54790/rccs.93>
- Castillo-Eslava, F., Mougán, C., Romero-Eche, A. y Staab, S. (2023). The Role of Large Language Models in the Recognition of Territorial Sovereignty. *arXiv*. <https://doi.org/10.48550/arxiv.2304.06030>
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q. y Xie, X. (2024). A Survey on Evaluation of Large Language Models. *ACM Transactions On Intelligent Systems And Technology*, 15(3), 1–45. <https://doi.org/10.1145/3641289>
- Charles, D. D., Mogoutov, A. y Baumard, N. (2024). Towards Transparency: Exploring LLM Trainings Datasets through Visual Topic Modeling and Semantic Frame. *arXiv*. <https://doi.org/10.48550/arxiv.2406.06574>
- Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., et

- aL (2024). Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis. *Journal Of Medical Internet Research*, 26, e53164. <https://doi.org/10.2196/53164>
- Chiang, C. y Lee, H. (2024). Merging Facts, Crafting Fallacies: Evaluating the Contradictory Nature of Aggregated Factual Claims in Long-Form Generations. *arXiv*. <https://doi.org/10.48550/arXiv.2402.05629>
- Chiang, W., Zhang, L., Sheng, Y., Angelopoulos, A. N., Li, T., Li, D., Zhang, H., Zhu, B., Jordan, M., Gonzalez, J. E. y Stoica, I. (2024). Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. *arXiv*. <https://doi.org/10.48550/arXiv.2403.04132>
- Choenni, R. y Shutova, E. (2024). Self-Alignment: Improving Alignment of Cultural Values in LLMs via In-Context Learning. *arXiv*. <https://doi.org/10.48550/arXiv.2408.16482>
- Cuellar-Rodríguez, A. (2023). Epistemology and ontology in Science: the challenge of Artificial Intelligence. *Anales de la Real Academia Nacional de Farmacia*, 89(03), 379–386. <https://doi.org/10.53519/analesranf.2023.89.03.09>
- Dierckx, L., Lindén, C. y Opdehl, A. L. (2023). The Information Disorder Level (IDL) Index: A Human-Based Metric to Assess the Factuality of Machine-Generated Content. En D. Ceolin, T. Caselli y M. Tulin (eds.), *Disinformation in Open Online Media. MISDOOM 2023* (pp. 60–71). https://doi.org/10.1007/978-3-031-47896-3_5
- Duan, S. yi, X., Zhang, P., Lu, T., Xie, X. y Gu, N. (2023). Denevil: Towards Deciphering and Navigating the Ethical Values of Large Language Models via Instruction Learning. *arXiv*. <https://doi.org/10.48550/arXiv.2310.11053>
- Esteban, J. L. G. (2014). La transformación del ecosistema mediático español: el caso de eldiario.es. *Revista Mediterránea de Comunicación*, 5(2), 159. <https://doi.org/10.14198/medcom2014.5.2.10>
- Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methodos Revista de Ciencias Sociales*, 11(2), m231102a10. <https://doi.org/10.17502/mrcs.v11i2.710>
- Gabriel, S., Colikyilmaz, A., Jha, R., Choi, Y. y Gao, J. (2020). GO FIGURE: A meta evaluation of factuality in summarization. *arXiv*. <https://arxiv.org/abs/2010.12834>
- Gao, Y., Baptista-Hon, D. T. y Zhang, K. (2023). The inevitable transformation of medicine and research by large language models. *MedComm – Future Medicine*, 2(2). <https://doi.org/10.1002/mef2.49>
- García-Santamaría, J. V. y Pérez-Serrano, M. J. (2020). Grupos de comunicación en España: madurez y profundas transformaciones en un final de ciclo. *Palabra Clave*, 23(4), 1–32. <https://doi.org/10.5294/pacla.2020.23.4.5>
- Ghodratnama, S. y Zakershahrah, M. (2023). Adapting LLMs for Efficient, Personalized Information Retrieval: Methods and Implications. *arXiv*. <https://doi.org/10.48550/arXiv.2311.12287>
- Guo, Z., Jin, R., Liu, C., Huang, Y., Shi, D., Supryadi, Yu, L., Liu, Y., Li, J., Xiong, B. y Xiong, D. (2023). Evaluating Large Language Models: A Comprehensive Survey. *arXiv*. <https://doi.org/10.48550/arXiv.2310.19736>
- Hader-Shoval, D., Asraf, K., Mizrahi, Y., Haber, Y. y Elyoseph, Z. (2024). Assessing the Alignment of Large Language Models With Human Values for Mental Health Integration: Cross-Sectional Study Using Schwartz's Theory of Basic Values. *JMIR Mental Health*, 11, e55988. <https://doi.org/10.2196/55988>
- Harandizadeh, B., Salinas, A. y Morstatter, F. (2024). Risk and Response in Large Language Models: Evaluating Key Threat Categories. *arXiv*. <https://doi.org/10.48550/arXiv.2403.14988>
- Hida, R., Kaneko, M. y Okazaki, N. (2024). Social Bias Evaluation for Large Language Models Requires Prompt Variations. *arXiv*. <https://doi.org/10.48550/arXiv.2407.03129>
- Huang, L., et al. (2023). A Survey on Hallucination in Large Language Models. *ACM Transactions On Office Information Systems*. <https://doi.org/10.1145/3703155>
- Jiao, J., Afroogh, S., Xu, Y. y Phillips, C. (2024). Navigating LLM Ethics: Advancements, Challenges, and Future Directions. *arXiv*. <https://doi.org/10.48550/arXiv.2406.18841>
- Larrondo-Ureta, A. y Peña-Fernández, S. (2024). La formación de periodistas en la era de la inteligencia artificial: aproximaciones desde la epistemología de la comunicación. *Anuario ThinkEPI*, 18. <https://doi.org/10.3145/thinkepi.2024.e18a11>
- Liang, Y., Xiao, J., Gan, W. y Yu, P. S. (2024). Watermarking Techniques for Large Language Models: A Survey. *arXiv*. <https://doi.org/10.48550/arXiv.2409.00089>
- Lin, Y. y Chen, Y. (2023). LLM-Eval: Unified Multi-Dimensional Automatic Evaluation for Open-Domain Conversations with Large Language Models. *arXiv*. <https://doi.org/10.48550/arXiv.2305.13711>
- Lindemann, N. F. (2023). Sealed knowledges: A critical approach to the usage of LLMs as search engines. En *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 985–986). <https://doi.org/10.1145/3600211.3604737>
- Liu, G., Wang, X., Yuan, L., Chen, Y. y Peng, H. (2023a). Prudent Silence or Foolish Babble? Examining Large Language Models' Responses to the Unknown. *arXiv*. <https://doi.org/10.48550/arXiv.2311.09731>
- Liu, N. F., Zhang, T. y Liang, P. (2023b). Evaluating Verifiability in Generative Search Engines. *arXiv*. <https://doi.org/10.48550/arXiv.2304.09848>
- Liu, Y., Cao, J., Liu, C., Ding, K. y Jin, L. (2024). Datasets for Large Language Models: A Comprehensive Survey. *arXiv*. <https://doi.org/10.48550/arXiv.2402.18041>
- Luitse, D. y Denkena, W. (2021). The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI. *Big Data y Society*, 8(2). <https://doi.org/10.1177/20539517211047734>
- Masís-González, T. M. (2024). La metáfora del shoggoth en la inteligencia artificial. *Pensamiento Actual*, 24(43). <https://doi.org/10.15517/pa.v24i43.62864>
- Masud, S., Singh, S., Hangya, V., Fraser, A. y Chakraborty, T. (2024). Hate Personified: Investigating the role of LLMs in content moderation. *arXiv*. <https://doi.org/10.48550/arXiv.2410.02657>
- McGowan, A., Gui, Y., Dobbs, M., Shuster, S., Cotter, M., Selloni, A., et al. (2023). ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search. *Psychiatry Research*, 326, 115334. <https://doi.org/10.1016/j.psychres.2023.115334>
- Menick, J., Trebecz, M., Mikulík, V., Aslanides, J., Song, F., Chadwick, M., et al. (2022). Teaching language models to support answers with verified quotes. *arXiv*. <https://doi.org/10.48550/arXiv.2203.11147>
- Mihajlov-Prokopović, A., Ievtović, Z. y Jovanović, Z. (2019). Digital challenges of local media of the Nišava District. *CM: Communication and Media*, 14(46), 5–32. <https://doi.org/10.5937/cm14-24293>
- Mokander, J., Schuett, J., Kirk, H. R. y Floridi, L. (2023). Auditing large language models: a three-layered approach. *AI And Ethics*, 4(4),

1085–1115. <https://doi.org/10.1007/s43681-023-00289-2>

Morato, O. y Nunes, D. J. C. (2023). O uso do design comportamental nas plataformas tecnológicas e as iniciativas de sua regulamentação: Um estudo do modelo gancho. *Revista Justiça Do Direito*, 37(2). <https://doi.org/10.5335/rjd.v37i2.14961>

Muhlgay, D., Ram, O., Magar, I., Levine, Y., Ratner, N., Belinkov, Y., et al. (2023). Generating benchmarks for factuality evaluation of language models. *arXiv*. <https://doi.org/10.48550/arxiv.2307.06908>

Muñoz-Vela, J. M. (2024). Inteligencia artificial generativa. Desafíos para la propiedad intelectual. *Revista de Derecho de la UNED*, (33), 17–75. <https://doi.org/10.5944/rduned.33.2024.41924>

Negreira-Rey, M., Vázquez-Herrero, J. y López-García, X. (2023). No people, no news: News deserts and areas at risk in Spain. *Media and Communication*, 11(3), 126–138. <https://doi.org/10.17645/mac.v11i3.6727>

Novelli, C., Casolari, F., Hacker, P., Spedicato, G. y Floridi, L. (2024). Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity. *Computer Law y Security Review*, 55, 106066. <https://doi.org/10.1016/j.clsr.2024.106066>

Olsen, R. y Hess, K. (2024). "It's New to Us": Exploring Authentic Innovation in Local News Settings. *Media and Communication*, 12, 7444. <https://doi.org/10.17645/mac.7444>

Patil, R. y Gudivada, V. (2024). A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs). *Applied Sciences*, 14(5), 2074. <https://doi.org/10.3390/app14052074>

Peña-Fernández, S., Peña-Alonso, U. y Eizmendi-Iraola, M. (2023). El discurso de los periodistas sobre el impacto de la inteligencia artificial generativa en la desinformación. *Estudios Sobre el Mensaje Periodístico*, 29(4), 833–841. <https://doi.org/10.5209/esmp.88673>

Pezeshkpour, P. y Hruschka, E. (2023). Large Language Models Sensitivity to The Order of Options in Multiple-Choice Questions. *arXiv* (Cornell University). <https://doi.org/10.48550/arxiv.2308.11493>

Piedra-Alegria, J. (2023). Anotaciones iniciales para una reflexión ética sobre la regulación de la Inteligencia Artificial en la Unión Europea. *Revista de Derecho*, 28, e3264. <https://doi.org/10.22235/rd28.3264>

Pit, P., Ma, X., Conway, M., Chen, Q., Bailey, J., Pit, H., Keo, P., Diep, W. y Jiang, Y. (2024). Whose Side Are You On? Investigating the Political Stance of Large Language Models. *arXiv*. <https://doi.org/10.48550/arxiv.2403.13840>

Rastogi, C., Ribeiro, M. T., King, N., Nori, H. y Amershi, S. (2023). Supporting human-AI collaboration in auditing LLMs with LLMs. En *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 913–926. <https://doi.org/10.1145/3600211.3604712>

Richter, L., He, X., Minervini, P. y Kusner, M. J. (2024). An Auditing Test To Detect Behavioral Shift in Language Models. *arXiv*. <https://doi.org/10.48550/arxiv.2410.19406>

Rotaru, G., Anagnoste, S. y Dancu, V. (2024). How Artificial Intelligence Can Influence Elections: Analyzing the Large Language Models (LLMs) Political Bias. En *Proceedings of the International Conference on Business Excellence*, 18(1), 1882–1891. <https://doi.org/10.2478/picbe-2024-0158>

Röttger, P., Hofmann, V., Pyatkin, V., Hinck, M., Kirk, H. R., Schütze, H. y Hovy, D. (2024). Political Compass or Spinning Arrow? Towards More Meaningful Evaluations for Values and Opinions in Large Language Models. *arXiv*. <https://doi.org/10.48550/arxiv.2402.16786>

Sarkar, D. (2024). Navigating the Knowledge Sea: Planet-scale answer retrieval using LLMs. *arXiv*. <https://doi.org/10.48550/arxiv.2402.05318>

Schwartz, L. S., Link, K. E., Daneshjoui, R. y Cortés-Penfield, N. (2023). Black box warning: Large language models and the future of infectious diseases consultation. *Clinical Infectious Diseases*, 78(4), 860–866. <https://doi.org/10.1093/cid/ciad633>

Shao, M., Basit, A., Karri, R. y Shafique, M. (2024). Survey of different Large Language Model Architectures: Trends, Benchmarks, and Challenges. *IEEE Access*, 12, 188664–188706. <https://doi.org/10.1109/access.2024.3482107>

Silveira, S. A. (2020). Responsabilidade algorítmica, personalidade eletrônica e democracia. *Revista Eletrônica Internacional de Economia Política da Informação da Comunicação e da Cultura*, 22(2), 83–96. <https://periodicos.ufs.br/eptic/article/view/12021>

Sun, D. Q., Abzaliev, A., Kotek, H., Xiu, Z., Klein, C. y Williams, J. D. (2023). DELPHI: Data for Evaluating LLMs' Performance in Handling Controversial Issues. *arXiv*. <https://doi.org/10.48550/arxiv.2310.18130>

Tlaie, A. (2024). Exploring and Steering the Moral Compass of Large Language Models. *arXiv*. <https://doi.org/10.48550/arxiv.2405.17345>

Vara-Miguel, A., Sánchez-Blanco, C., Chalequer, C. S. S. y Negredo, S. (2021). Funding sustainable online news: Sources of revenue in digital-native and traditional media in Spain. *Sustainability*, 13(20), 11328. <https://doi.org/10.3390/su132011328>

Wan, A., Wallace, E. y Klein, D. (2024). What evidence do language

PUBLICIDAD | ADVERTISING | PUBLICITAT

OCM Observatorio de Cybermedios

<https://www.upf.edu/web/ocm>

El Observatorio de Cybermedios (OCM) es una producción del Grupo de Investigación en Documentación Digital y Comunicación Interactiva (DigiDoc) del Departamento de Comunicación de la Universitat Pompeu Fabra.

El OCM forma parte del proyecto del Plan Nacional "Parámetros y estrategias para incrementar la relevancia de los medios y la comunicación digital en la sociedad: curación, visualización y visibilidad (CUMCOM)". PID2021-123579GB-100 (MICINN), Ministerio de Ciencia e Innovación (España).



Universitat
Pompeu Fabra
Barcelona

Departament d'Informació
i Comunicació



Ministerio de
Ciencia e Innovación



Unió Europea

models find convincing? *arXiv*. <https://doi.org/10.48550/arXiv.2402.11782>

Wang, Y., Zhong, W., Li, L., Mi, F., Zeng, X., Huang, W., Shang, L., Jiang, X. y Liu, Q. (2023). Aligning Large Language Models with Human: A Survey. *arXiv*. <https://doi.org/10.48550/arXiv.2307.12966>

Wu, K., Wu, E., Cassasola, A., Zhang, A., Wei, K., Nguyen, T., et al. (2024). How well do LLMs cite relevant medical references? An evaluation framework and analyses. *arXiv*. <https://doi.org/10.48550/arXiv.2402.02008>

Yue, X., Wang, B., Zhang, K., Chen, Z., Su, Y. y Sun, H. (2023). Automatic Evaluation of Attribution by Large Language Models. *arXiv*. <https://doi.org/10.48550/arXiv.2305.06311>

Zhang, D., et al. (2024). A survey of datasets in medicine for large language models. *Intelligence y Robotics*, 4(4), 457–478. <https://doi.org/10.20517/ir.2024.27>

Zhao, H., et al. (2024). Towards uncovering how large language model works: An explainability perspective. *arXiv*. <https://doi.org/10.48550/arXiv.2402.10688>

Zheng, X., Wang, L., Liu, Y., Ma, X., Shen, C. y Wang, C. (2025). CALM: Curiosity-Driven Auditing for Large Language Models. *arXiv*. <https://doi.org/10.48550/arXiv.2501.02997>

Zhu, Y., Yuan, H., Wang, S., Liu, J., Liu, W., Deng, C., Dou, Z. y Wen, J. (2023). Large Language Models for Information Retrieval: A Survey. *arXiv*. <https://doi.org/10.48550/arXiv.2308.07107>

CV

Barbara Sarriónandía

- bsarrionandia001@ikasle.ehu.eus
- <https://orcid.org/0009-0006-0053-3730>
- Es licenciada en Periodismo por la Universidad de Navarra y obtuvo el Máster Universitario en Formación del Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanzas de Idiomas, con especialidad en Geografía e Historia, por la Universidad Internacional de La Rioja (UNIR). Su investigación se centra en la Inteligencia artificial, la desinformación y la alfabetización mediática. A lo largo de dos décadas ha desarrollado su trayectoria profesional en medios de comunicación e Instituciones tanto a nivel estatal como internacional.

Simón Peña-Fernández

- simon.penaf@ehu.eus
- <https://orcid.org/0000-0003-2080-3241>
- Profesor Pleno en el Departamento de Periodismo de la Universidad del País Vasco (UPV/EHU). Sus principales líneas de investigación son el ciberperiodismo, la comunicación digital y la Inteligencia artificial. Es el investigador principal, junto con Koldobika Meso, del proyecto de Investigación "Impacto de la Inteligencia artificial y los algoritmos en los cibermedios, los profesionales y las audiencias" (PID2022-138391GB-I00), financiado por el Ministerio de Ciencia e Innovación.

Jesús Ángel Pérez-Dasilva

- jesusangel.perez@ehu.eus
- <https://orcid.org/0000-0002-3383-4859>
- Profesor Pleno en el Departamento de Periodismo de la Universidad del País Vasco (UPV/EHU). Sus principales líneas de investigación son el ciberperiodismo, la comunicación social, las redes sociales y la innovación social. Es el investigador principal del proyecto "Impacto de la Inteligencia artificial en los medios de comunicación vascos y sus profesionales" (US 23/10), financiado por la Universidad del País Vasco (UPV/EHU).

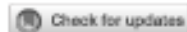
PUBLICIDAD | ADVERTISING | PUBLICITAT

HTXT
n. 30 / 2025

**Llegar a todas las audiencias:
cuando el periodismo se
enfrenta a la desertización
informativa**

María Cruz Negrera Rey
Universidade de Santiago de Compostela

<https://doi.org/10.31009/hipertext.net.2025.30.12>



OPEN ACCESS

EDITED BY
Maria O'Brien,
University of Galway, Ireland

REVIEWED BY
Michał Kuś,
University of Wrocław, Poland

*CORRESPONDENCE
Simón Peña-Fernández
✉ simon.pena@ehu.es

RECEIVED 02 December 2024
ACCEPTED 20 May 2025
PUBLISHED 09 June 2025

CITATION
Sarrionandia B, Peña-Fernández S, Ángel
Pérez-Dasilva J and Larrondo-Ureta A (2025)
Artificial intelligence training in media:
addressing technical and ethical challenges
for journalists and media professionals.
Front. Commun. 10:1537918.
doi: 10.3389/fcomm.2025.1537918

COPYRIGHT
© 2025 Sarrionandia, Peña-Fernández,
Ángel Pérez-Dasilva and Larrondo-Ureta. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Artificial intelligence training in media: addressing technical and ethical challenges for journalists and media professionals

Barbara Sarrionandia, Simón Peña-Fernández*,
Jesús Ángel Pérez-Dasilva and Ainara Larrondo-Ureta

Department of Journalism, University of the Basque Country (UPV/EHU), Bilbao, Spain

The rise of Artificial Intelligence (AI) is presenting both technical and ethical challenges for media organisations, creating an urgent need for professional training. This study explores how media professionals in the Basque Country are equipping themselves to face these challenges. Using a mixed-method approach, it combines a survey of 504 active professionals with in-depth interviews with six innovation leaders from major regional media outlets. The findings reveal that only 14.1% of professionals have undergone AI training, mostly through self-learning. Larger, internationally focused companies are more proactive in providing training, while local and traditional media organisations show significant gaps. Technical and managerial roles are leading the way in adopting AI, whereas newsroom staff are notably behind. The study highlights the pressing need to enhance AI training, with a particular focus on ethical and technical aspects, both through in-house programmes and formal education pathways.

KEYWORDS

artificial intelligence, journalists, media, training, digital transformation

1 Introduction

The adoption of AI in newsrooms and media organisations has introduced a range of complex challenges for journalists and media professionals (Beckett et al., 2023; Coeckelbergh, 2020). As news production and distribution evolve, AI offers opportunities for efficiency and innovation, but it also brings risks and vulnerabilities (Brennen et al., 2018).

Much like previous technological advancements in the digitalisation of media, AI is a tool that requires careful consideration of its ethical implications. Challenges such as reinforcing biases through technology (Leiser, 2022), increasing misinformation and manipulation (García-Marín and Salvat-Martínez, 2022), and safeguarding privacy, can all be mitigated with robust media literacy and targeted training. These measures empower journalists to not only use AI tools effectively but also navigate the ethical dilemmas they present (Aissani et al., 2023; Baldessar and Zandoménico, 2024; Barceló-Ugarte et al., 2021; Helberger and Diakopoulos, 2023; Ufarte Ruiz et al., 2021).

AI, with its continuous evolution, holds the potential to transform our lives, work, and relationships. However, it also raises serious ethical and social justice concerns, particularly regarding the widening of existing disparities (e.g., race and gender) (Brundage et al., 2018) and the growing sophistication and dissemination of misinformation, especially through advanced deepfake creation (Phelan, 2022; Sohrawardi et al., 2024).

Nonetheless, AI is not solely a cause for concern among journalism professionals (Peña-Fernández et al., 2023a). In its more constructive application, it can serve as a powerful tool to combat informational disorders and malicious content (Bontridder and Pouillet, 2021;

Manfredi-Sánchez and Ufarte-Ruiz, 2020; Rubin, 2022). This technology has the ability to match the speed and sophistication of digital falsehoods, reducing the effort and time required by fact-checking professionals and enhancing their capacity to respond effectively to misinformation.

For AI to strengthen journalism's role in supporting democracy (Lin and Lewis, 2022), it is vital to establish adequate regulations. These should include clear guidelines on authorship (Hofeditz et al., 2021; Díaz-Noci et al., 2024) and enforce transparency (Larsson and Heintz, 2020). AI's integration into media does not only impact editorial workflows but also necessitates significant shifts in the skills and competencies required for journalists (Danzon-Chambrud and Cornia, 2023; Demmar and Neff, 2023; Gómez-Diogo, 2022; Lopezosa et al., 2023; Noain Sánchez, 2022; Ufarte-Ruiz et al., 2020).

2 State of the art

The prominence of AI, heightened since late 2022 with the emergence of generative AI and tools like ChatGPT, has spurred significant innovation across various fields, including journalism (Bakke and Barland, 2022; García-Orosa et al., 2023; Tejedor and Vila, 2021). Although AI is not a novel concept in media (Manfredi-Sánchez and Ufarte-Ruiz, 2020), efforts to automate news production by major outlets and agencies have provided examples over the past two decades. For instance, The Big Ten Network has used automated systems for sports reporting since 2007, while Quakebot at The Los Angeles Times has been offering real-time earthquake updates since 2014. Agencies such as Reuters and Associated Press have also automated parts of their services (Danzon-Chambrud, 2021; Sánchez-García et al., 2023).

A revolutionary shift has been the democratisation of generative AI tools, with ChatGPT as a leading example. Numerous companies have introduced similar platforms (Motlagh et al., 2023; Singh et al., 2023). The societal adoption of these technologies has been remarkably swift, with ChatGPT reaching 100 million users within just 2 months, outpacing TikTok (9 months) and Instagram (two and a half years) (Milmo, 2023).

The emergence of such tools can be seen as the latest phase in media's digital transformation, redefining its nature within the framework of the Fourth Industrial Revolution (Micó et al., 2022; Dhiman, 2023). This transformation is expected to have a widespread impact across all areas of communication, from news production processes (Sánchez-García et al., 2023) to adjustments in normative values (Peña-Fernández et al., 2023b).

AI is reshaping journalism in multiple ways. By automating content production and reducing reporters' workload, it fosters innovation and professional specialisation, allowing journalists to focus on more cognitive and creative tasks (Wu et al., 2019). Generative AI can also enable coverage of previously unprofitable topics (Atasoy et al., 2021) and deliver more personalised content (Hermann, 2022).

Integrating AI into newsrooms seems inevitable, as this technology has become essential for the evolution of both media and journalism itself (Terol, 2023; García-Orosa et al., 2023; Gutiérrez-Caneda et al., 2023).

For professionals, concerns about AI extend beyond industrial or sectoral issues such as changes in production processes, potential job

losses, or shifts in required skill sets. There is considerable uncertainty about its impact on public opinion and democratic societies.

Since its release in November 2022, OpenAI's platform has raised concerns regarding the spread of misinformation (Aydm and Karaarslan, 2023; Opdahl et al., 2023). Although the March 2023 release of ChatGPT-4 showed improvements in source citation, misinformation remains a persistent issue (Gutiérrez-Caneda et al., 2023).

During this period, discourse among media professionals has increasingly focused on the potential negative consequences of AI in journalism (Beckett et al., 2023). Concerns include the potential for AI reliance to degrade journalistic quality (Calvo-Rubio and Rojas-Torrijos, 2024), reinforce inherent biases (Cloudy et al., 2023; Leiser, 2022), and encourage unethical practices such as content farms and plagiarism (Palacios Tapia, 2023; Subiela-Hernández and Vizcaino-Laorga, 2023).

Despite these risks, the primary concern among professionals since the advent of generative AI remains its role in spreading false information, fostering misinformation, and exacerbating polarisation (Berrocal-Gonzalo et al., 2023; Peña-Fernández et al., 2023a).

In response to the growing social and political concerns about misinformation, governments and institutions have intensified efforts to mitigate its impact. Recent measures include the development of tools like fact-checking algorithms, detection systems (e.g., random forests), bots and chatbots, and monitoring platforms, all of which heavily rely on AI (Arias Jiménez et al., 2022; Garriga et al., 2024; Moreno Espinosa et al., 2024; Alonso González and Sánchez Gonzales, 2024). These initiatives combine information verification with media literacy programmes, ongoing training, and professional development for journalists.

Given the technical and ethical complexities involved, training for current and future professionals has become a critical priority, particularly in a context where technological evolution is outpacing existing educational structures.

In this context, this study examines how media professionals in the Basque Country perceive the opportunities and risks associated with AI, their current training levels, and what they consider necessary to adapt to this rapidly evolving landscape.

3 Methodology

A mixed-method approach was employed, combining quantitative survey data with qualitative insights gathered through semi-structured in-depth interviews. This methodology was chosen to provide a comprehensive understanding of journalists' and media workers' experiences and perceptions while also quantifying trends and relationships across variables.

The survey was conducted online, with supplementary telephone support, during May and June 2024. A total of 504 responses were collected from journalists and communication sector professionals working in media, companies, or institutions in the Basque Country. Participants were identified through the Open Communication Guide provided by the Basque Government,¹ which lists active media

¹ <https://gida.irekia.euskadi.eus>

organisations in the region, and through the Basque Journalists' Association.

According to existing data (Basque Government, 2022; Pérez et al., 2023), it is estimated that approximately 5,000 people work in the media sector in Euskadi. Therefore, for a 95% confidence level, the margin of error of the survey is $\pm 4.15\%$.

The survey employed an online panel and ensured respondent anonymity. No personal data were collected, eliminating risks related to data storage and handling. The sample included 276 men (55.1%), 223 women (44.5%), and two individuals who identified as "other" (0.4%). Variables such as years of professional experience, type of media, scope of the organisation, job role, and responsibility levels were also analysed.

In addition to the survey, six semi-structured in-depth interviews were conducted with senior figures specialising in technology, innovation, and AI from the most prominent media outlets based in the Basque Country. These interviews, lasting between 30 and 50 min, were conducted via videoconference in April and May 2024.

Code	Media	Media Type	Position
I1	Vocento/El Correo	Newspaper	Head of New Technological Trends and Quality
I2	EiTB	Public broadcasting service	Deputy Director of EiTB Media
I3	Noticias Group/Diario	Newspaper	Director of the Technology Department
I4	Natz	Newspaper	Content and Forecast Manager
I5	Berria	Newspaper	Head of Information Technologies
I6	Cadena SER/ Radio Bidasoa	Radio	Chief Editor

Source: Own elaboration.

Based on the state of the art, this study seeks to answer the following research questions:

- RQ1: What specific AI training have journalists received so far?
- RQ2: What are the perceived training needs among journalists in this area?
- RQ3: What are the current policies of media organisations regarding AI training, according to media leaders?

4 Results and analysis

4.1 AI training among media professionals

The findings reveal that AI training levels among journalists are generally low. Despite the growing importance of this technology in the media industry, only 14.1% of professionals have received any form of training in this area. In terms of gender, women (15.9%) reported slightly higher training rates than men (12.7%).

Significant differences in AI training are observed across different types of media (Table 1). For instance, professionals engaged in roles further from traditional journalism, such as corporate communication and advertising, show a stronger commitment to AI training.

Among respondents, 44.4% of those working in advertising agencies and 21.2% of those in communication departments reported receiving AI training, far exceeding the levels observed among those in purely journalistic roles. A more business-oriented and economically driven approach to communication activities may explain this higher investment in AI training, as it offers opportunities to optimise both time and costs in content production.

A notable difference in AI training levels is also evident among professionals working in outlets that produce purely journalistic content. Those employed in digital-native sections or outlets reported higher training rates (25.8%) compared to their counterparts in more traditional media formats, such as radio (10.8%), print (9.6%), and television (5.8%).

In summary, the closer media professionals are to traditional journalistic tasks and outlets, the lower their levels of AI training tend to be. Conversely, training is significantly higher among individuals working in digital areas of media organisations. Similarly, the further removed their roles are from core journalistic activities and the closer they are to persuasive functions—such as advertising or corporate communication—the greater the likelihood of having received AI training.

The findings also highlight a correlation between the size and scope of media organisations and the AI training their employees receive (Table 1). Larger organisations with broader dissemination areas show higher training levels, with international outlets reporting nearly double the training rate (21.7%) of local outlets (12.2%). These results suggest that the adoption of AI in media may exacerbate the existing technological gap that has emerged since the onset of media digitalisation. Smaller outlets, in particular, are less prepared to meet this challenge.

The data further confirms that professionals more distanced from traditional journalistic tasks, such as editors and presenters, have received less AI training compared to those in technical roles or working in corporate communication and advertising (Table 2).

Moreover, it is unsurprising that individuals in leadership roles have, on average, received twice as much training as those without such responsibilities (19.9% compared to 9.4%). The disruptive nature of AI technology, its potential impact on work processes, and the availability of greater training opportunities are likely factors influencing this trend. This pattern suggests that more experienced professionals, particularly those in managerial or strategic positions, are more engaged in adopting and implementing AI technologies. Their roles often require staying updated on technological advancements to make informed strategic decisions.

Similarly, employees with longer tenures in their organisations are more likely to have received training. Their stability and seniority in the workplace, coupled with their presence in leadership roles, likely contribute to these higher training levels. Additionally, a perceived technological gap may have driven them to seek further training. These findings underline the importance of integrating AI training into the early and mid-stages of professional development to equip workers for technological challenges.

TABLE 1 Professionals who have received AI training, by media type.

	Men		Women		Total	
	N	%	N	%	N	%
Type of media						
Press	14	12,2	5	6,1	19	9,6
Radio	6	11,1	3	10,3	9	10,8
TV	2	5,7	2	5,9	4	5,8
Digital	6	17,6	11	34,4	17	25,8
Press Office	1	5,0	12	29,3	13	21,3
Advert. Agency	5	45,5	3	42,9	8	44,4
Broadcasting area						
Local	8	22,9	2	4,3	10	12,2
Regional	21	10,9	23	17,7	44	13,6
National	6	16,2	6	15,8	12	16,0
International	0	0,0	5	41,7	5	21,7
Total	35	12,7	36	15,9	71	14,1

Source: Own elaboration.

TABLE 2 Professionals who have received AI training, by job position.

	Men		Women		Total	
	N	%	N	%	N	%
Job type						
Editor	22	13,5	21	14,1	43	13,8
Presenter, host	3	4,7	3	9,7	6	6,3
Graphic (design, photographer)	1	11,1	2	28,6	3	18,8
Technical (cameraman, sound)	2	100	0	0	2	66,7
Producer, webmaster	0	0	3	30,0	3	15,8
Management	3	23,1	2	28,6	5	25,0
Public relations	3	27,3	3	21,4	6	24,0
Copywriter, creative	0	0	1	25,0	1	20,0
Other	1	25,0	1	33,3	2	28,6
Experience						
Less than 5 years	3	5,7	6	12,0	9	8,7
From 5 to 10 years	7	15,9	4	14,8	11	15,5
From 10 to 20 years	2	3,9	6	10,5	8	7,4
More than 20 years	23	18,0	20	21,7	43	19,5
Responsibility role						
Yes	22	17,9	23	22,3	45	19,9
No	13	8,5	13	10,6	26	9,4
Total	35	12,7	36	15,9	71	14,1

Source: Own elaboration.

4.2 AI training policies in media organisations

AI training is a critical factor for adopting this technology effectively. Among respondents who reported never using AI, 43.3% cited a lack of training as the primary barrier.

The study found that only 8.5% of participants received AI training provided by their organisations, whether online or in person, compared to 12.9% who pursued training independently.

By organisation type, advertising agencies, digital media, and communication departments stand out for their commitment to AI training (Table 3). In contrast, traditional and journalistic

TABLE 3 Type of AI training received, by media type.

	Self online course		Self face-to-face course		Self-study		Company online course		Company face-to-face course	
	N	%	N	%	N	%	N	%	N	%
Type of media										
Press	5	2,5	2	1,0	17	8,5	4	2,0	8	4,0
Radio	5	6,0	2	2,4	7	8,4	3	3,6	2	2,4
TV	2	2,9	0	0,0	4	5,8	0	0,0	3	4,3
Digital	13	19,7	5	7,6	14	21,2	8	12,1	8	12,1
Press Office	6	9,8	2	3,3	10	16,4	7	11,5	4	6,6
Advert. Agency	2	11,1	1	5,6	7	38,9	6	33,3	4	22,2
Broadcasting area										
Local	4	4,9	3	3,7	8	9,8	3	3,7	5	6,1
Regional	21	6,5	6	1,9	37	11,4	15	4,6	16	4,9
National	7	9,3	2	2,7	11	14,7	8	10,7	5	6,7
International	2	8,7	1	4,3	4	17,4	2	8,7	3	13,0
Total	34	6,7	12	2,4	28	11,9	29	5,6	60	5,8

Source: Own elaboration.

TABLE 4 Type of AI training received, by job position.

	Self online course		Self face-to-face course		Self-study		Company online course		Company face-to-face course	
	N	%	N	%	N	%	N	%	N	%
Job type										
Editor	21	6,7	7	2,2	38	12,1	16	5,1	17	5,4
Presenter, host	3	3,2	1	1,1	5	5,3	2	2,1	2	2,1
Graphic (design, photographer)	2	12,5	0	0,0	3	18,8	0	0,0	0	0,0
Technical (cameraman, sound)	0	0,0	0	0,0	2	66,7	0	0,0	1	33,3
Producer, webmaster	1	5,3	1	5,3	2	10,5	0	0,0	1	5,3
Management	2	10,0	1	5,0	4	20,0	3	15,0	3	15,0
Public relations	3	12,0	2	8,0	4	16,0	5	20,0	4	16,0
Copywriter, creative	0	0,0	0	0,0	1	16,7	1	16,7	1	16,7
Other	2	28,6	0	0,0	1	14,3	1	14,3	0	0,0
Experience										
Less than 5 years	6	5,8	1	1,0	9	8,7	2	1,9	2	1,9
From 5 to 10 years	4	5,6	2	2,8	9	12,7	4	5,6	5	7,0
From 10 to 20 years	3	2,8	2	1,9	7	6,5	3	2,8	1	0,9
More than 20 years	21	9,5	7	3,2	35	15,8	19	8,6	21	9,5
Responsibility role										
Yes	21	9,2	7	3,1	36	15,8	19	8,3	19	8,3
No	13	4,7	5	1,8	24	8,7	9	3,3	10	3,6
Total	34	6,7	12	2,4	60	11,9	28	5,6	29	5,8

Source: Own elaboration.

media outlets have offered little training in this area to their employees.

This trend aligns with the types of roles that reported higher training levels (Table 4), which were predominantly technical and

commercial positions—areas more peripheral to the production of current affairs content.

In terms of the scope of media operations, 13% of employees at international outlets reported seeking AI training independently, the

highest percentage among all categories. This reflects a greater personal initiative to acquire AI skills in broader dissemination environments. However, despite the correlation between media reach and AI training, a significant proportion of professionals across all levels remain untrained. In local outlets, 87.8% of professionals reported no AI training, a figure that, while slightly lower, remains substantial in regional (86.4%), national (84%), and international (78.3%) contexts.

Technical professionals stand out as the most trained group in AI, with 66.7% having received training. This aligns with the technical nature of AI, which demands expertise in data analysis, programming, and advanced software use. In management and leadership roles, 25% of professionals reported AI training, highlighting its strategic importance for decision-making and innovation.

Media leaders, however, emphasise that an exclusively technical approach to AI is insufficient. They stress the need to integrate technical understanding with ethical considerations to address the core challenges of journalism. As one interviewee noted: "We need people in newsrooms trained in these topics, and, of course, at a technical level. But perhaps, also on an ethical level, it is important to have technical knowledge of AI [...] to know when and how to use it" (14).

Following technical roles, graphic design professionals (18.75%) and production staff (15.8%) reported higher training levels. Advertising and corporate communication professionals also showed notable rates (16.7 and 14%, respectively). Journalists in editorial roles (13.7%) and presenters (6.3%) had the lowest training rates. According to media leaders, the limited immediate application of AI in fast-paced editorial work may explain this lower adoption. As one leader remarked: "Internally, it has not affected us yet; at the moment, we are not using anything. But given the breadth of what we cover, it could be useful, yes. Something has indeed been considered" (16).

The analysis also reveals a strong overlap between individual initiative and organisational support for training. Among those who received any form of training, 52.1% did so through both personal and company efforts, 39.4% relied solely on self-directed learning, and only 8.5% received training exclusively from their employer. This suggests that corporate training initiatives largely benefit those already motivated to learn, leaving a majority without access to such opportunities. Traditional media outlets and more journalistic roles appear to have the most room for improvement in this regard.

This self-motivated approach aligns with comments from media leaders: "It does not all depend on companies; individuals need that drive to keep training and equip themselves with more tools to do their job. Continuous training is not just a company's responsibility; it is also part of the worker's responsibility" (16).

The findings underscore the need for enhanced corporate training opportunities to provide media professionals with a deeper and more practical understanding of AI. Media leaders emphasise that this training must be ongoing: "After just 2 months, you are already outdated" (11). Thus, training efforts cannot be one-off initiatives but must be sustained over time: "We need to keep updating our skills because technology evolves so rapidly. What was cutting-edge a year ago is likely obsolete today" (12).

Given the rapid pace of change in AI, the focus has shifted to continuous professional development. As one leader noted: "The market demands far ahead of what educational institutions can

offer. These are skills everyone working with AI should know. For example, you must be careful about what you provide to AI systems, as you could be sharing confidential internal company information" (11).

While media leaders recognise the urgency of training to address AI-driven changes, they also stress the importance of foundational knowledge in technical aspects of AI (12, 14, 15). One leader commented: "We need to improve competencies and prioritise training for technology use in general, as media organisations include teams of all ages with varying technological capacities" (13).

5 Discussion and conclusions

This study on AI training among communication professionals in the Basque Country reveals a significant gap in specialised training, despite the growing importance of AI in the media industry (Husnain et al., 2024). The findings indicate that only a small percentage of journalists have received AI training so far, most of which was self-directed. The continuous training policies offered by companies and the self-directed learning initiatives of journalists emerge as essential needs in an environment undergoing rapid and profound transformation.

The disparities identified among different types of organisations and professional profiles highlight key areas where training efforts should be directed. Larger communication companies with broader dissemination reach have made greater investments in training their professionals, while smaller, local organisations have so far made fewer efforts in this area. This disparity suggests that AI could further widen the technological gap already created by media digitalisation, intensifying the divide between major international platforms and smaller, resource-constrained outlets.

Similarly, the data also reveal a differential adoption of AI between core and peripheral aspects of media operations. While areas adjacent to journalistic activity (technical, commercial, etc.) are adopting this technology earlier through training, professionals directly involved in content creation are lagging behind in terms of training received. This may be influenced by scepticism among journalists about the potential impact of AI on their profession (Peña-Fernández et al., 2023b).

Media leaders in the Basque Country acknowledge the need for greater ongoing training in AI (Noain Sánchez, 2022). Technological training is viewed as essential, not only to enhance individual skills but also to ensure the responsible and effective use of AI in media. In the short term, organisations are advised to provide in-house or company-supported training programmes to enable collaboration between media professionals and technical staff, fostering awareness of the risks associated with improper AI use. In the longer term, university-level training programmes should also address the broader implications of AI applications in media, focusing on its potential to generate misinformation, perpetuate biases, and exacerbate social inequalities (Opdahl et al., 2023).

In summary, this research highlights the urgent need to expand AI literacy and training among communication professionals (Deuze and Beckett, 2022; Gómez-Diago, 2022; Larrondo-Ureta and Peña-Fernández, 2024). Only through continuous and comprehensive AI training—encompassing both technical and ethical aspects—can media professionals effectively meet contemporary challenges and ensure the responsible use of AI in the media industry.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

BS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Writing – original draft. SP-F: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing – review & editing. JP-D: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. AL-U: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This article is part of the

research project “Impact of artificial intelligence and algorithms on online media, journalists and audiences” (PID2022-138391OB-I00) funded by the Spanish Ministry of Science, Innovation and Universities and by the European Commission NextGeneration EU/PRTR and “The Impact of Artificial Intelligence on Basque Media and Their Professionals” (US 23/10), funded by the University of the Basque Country (UPV/EHU). The authors belong to the Basque University’s Consolidated Research Group System “Gureiker” (IT-1496-22).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. Generative AI was used to adapt references to Harvard format and to check English language.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Atouani, R., Abdallah, R. A. Q., Taha, S., and Al Adwain, M. N. (2023). “Artificial intelligence tools in media and journalism: roles and concerns.” in *Proceedings of the 2023 International Conference on Multimedia Computing, Networking and Applications (MCNA)*, June, pp. 19–26. IEEE.
- Alonso González, M., and Sánchez González, M. (2024). Inteligencia artificial en la verificación de la información política: Herramientas y tipología. *Más Poder Local* 56, 27–45. doi: 10.56151/mas poder local.215
- Arias Jiménez, B., Rodríguez-Hidalgo, C., Mier-Sanmartín, C., and Coronel-Salas, G. (2022). “Use of chatbots for news verification: Communication and applied technologies.” in *Proceedings of ICDMTA 2022 Singapore* Springer Nature Singapore, pp. 133–143.
- Atalay, B., Hic, M., and Tatal, V. (2021). Towards the artificial intelligence management in sports. *Int. J. Sports Exer. Train. Sci. - IJSETS*, 7, 100–113. doi: 10.18826/ijsets.7.4.5994
- Aydın, Ö., and Karadeniz, E. (2023). Is ChatGPT leading generative AI? What is beyond expectations? *Acad. Platform J. Eng. Smart Syst.* 11, 118–134. doi: 10.21541/apjss.1293702
- Bakke, N. A., and Barland, J. (2022). Disruptive innovations and paradigm shifts in journalism as a business: from advertisers first to readers first and traditional operational models to the AI factory. *SAGE Open* 12, 1–13. doi: 10.1177/21582440221094819
- Baldassar, M. J., and Zandomeni, R. (2024). “Journalistic ethics in the face of news produced by artificial intelligence” in *Creating culture through media and communication*, eds. S. V. Moreira, K. Moles, L. Robinson and J. Schulz, (London: Emerald Publishing Limited), 131–137.
- Barcidó-Ugarte, T., Pérez-Tornero, J. M., and Vila-Fumika, P. (2021). Ethical challenges in incorporating artificial intelligence into newsrooms. *News Media Innovation Reconsidered: Ethics and Values in a Creative Reconstruction of Journalism*, 138–153.
- Basque Government (2022). Censo del mercado de trabajo. Specific Statistical Body of the Department of Labor and Employment. Available online at: https://www.eustat.eus/censo-del-mercado-de-trabajo-oferta/inf0021420_c.pdf (Accessed March 22, 2024).
- Beckett, C., Sangatelli, P., and Palomo, R. (2023). “New frontiers of intelligent journalism” in *Blurring boundaries of journalism in digital media: New actors, models and practices*, eds. M. C. Negreiros-Rex, J. Vázquez-Herrero, J. Stato-García and X. López-García, (Cham: Springer International Publishing), 275–288.
- Berrocal-Gonzalo, S., Watabord, S., and Gómez-García, S. (2023). Polarización política y medios de comunicación, su impacto en la democracia y en la sociedad. *Prof. Inform.* 32:e320622. doi: 10.3145/ept.2023.nov.22
- Bontridder, N., and Poulet, Y. (2021). The role of artificial intelligence in disinformation. *Data Policy* 3:e32. doi: 10.1017/dap.2021.20
- Brennen, J. S., Howard, P. N., and Nielsen, R. K. (2018). An industry-led debate: how UK media cover artificial intelligence. Reuters Institute for the Study of Journalism. Available online at: <https://www.oxfordmartin.ox.ac.uk/publications/an-industry-led-debate-how-uk-media-cover-artificial-intelligence> (Accessed March 22, 2024).
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., et al. (2018). The malicious use of artificial intelligence: forecasting, prevention, and mitigation. *arXiv*. Available online at: <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf> (Accessed March 22, 2024).
- Calvo-Rubio, L. M., and Rojas-Torrijos, J. L. (2024). Criteria for journalistic quality in the use of artificial intelligence. *Commun. Soc.* 37, 247–259. doi: 10.15581/003.37.2.247-259
- Cloudy, J., Banks, J., and Bowman, N. D. (2023). The str(AI)ght scoop: artificial intelligence cues reduce perceptions of hostile media bias. *Digit. Journal.* 11, 1577–1596. doi: 10.1080/21670811.2021.1969974
- Cockelbergh, M. (2020). AI ethics. Cambridge: MIT Press.

- Danzon-Chambaud, S. (2021). A systematic review of automated journalism scholarship: guidelines and suggestions for future research. *Open Res. Europe* 1. doi: 10.12688/openres.13096.1
- Danzon-Chambaud, S., and Cornia, A. (2023). Changing or reinforcing the "rules of the game": a field theory perspective on the impacts of automated journalism on media practitioners. *Journal. Pract.* 17, 174–188. doi: 10.1080/17512786.2021.1919179
- Demm, K., and Neff, T. (2023). Generative AI in journalism education: mapping the state of an emerging space of concerns, opportunities, and strategies. *Journal. Educ.* 12, 47–58.
- Deuze, M., and Beckett, C. (2022). Imagination, algorithms and news: developing AI literacy for journalists. *Digit. J.* 10, 1913–1918. doi: 10.1080/21670811.2022.2119152
- Dhiran, D. B. (2023). The challenges of computational journalism in the 21st century. SSRN. Available online at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4456378 (Accessed March 22, 2024).
- Díaz-Noci, J., Peña-Fernández, S., Mesa-Ayerdí, K., and Larrondo-Ureta, A. (2024). The influence of AI in the media workforce: how companies use an array of legal remedies. *Triposos* 55, 33–54. doi: 10.51698/triposos.2024.55.03
- García-Orosa, B., Canavilhas, J., and Vázquez-Herrero, J. (2023). Algorithms and communication: A systematic literature review. *Comunicar* 74, 9–21. doi: 10.3916/C74-2023-01
- García-Marin, D., and Salvat-Martínez, G. (2022). Tendencias en la producción científica sobre desinformación en España. *Revista sistemática de la literatura* (2016–2021). *adComunica* 23, 23–50. doi: 10.6035/adcomunica.6045
- Garriga, M., Ruiz-Incertis, R., and Magallón-Rosa, R. (2024). Artificial intelligence, disinformation and media literacy proposals around deepfakes. *Observatorio (OBS)**, 18.
- Gómez-Díaz, G. (2022). Perspectivas para abordar la inteligencia artificial en la enseñanza de periodismo: Una revisión de experiencias investigadoras y docentes. *Rev. Lat. Comun. Soc.* 80, 29–46. doi: 10.4185/RICS-2022-1542
- Gutiérrez-Caneda, B., Vázquez-Herrero, J., and López-García, X. (2023). AI application in journalism: chatGPT and the uses and risks of an emergent technology. *Prof. Inform.* 32:e320514. doi: 10.3145/pt.2023.sep.14
- Helberger, N., and Dikopoulos, N. (2023). The European AI act and how it matters for research into AI in media and journalism. *Digit. Journal* 11, 1751–1760. doi: 10.1080/21670811.2022.2082505
- Hermann, E. (2022). Artificial intelligence and mass personalization of communication content: an ethical and literacy perspective. *New Media Soc.* 24, 1258–1277. doi: 10.1177/14614448211022702
- Hofstede, L., Mihalache, M., Hofstede, J., and Stegla, S. (2021). Do you trust an AI journalist? A credibility analysis of news content with AI authorship. In: *ECIS Proceedings*, June.
- Huaini, M., Imran, A., and Tarcon, H. K. (2024). Artificial intelligence in journalism: examining prospectus and obstacles for students in the domain of media. *J. Asian Dev. Stud.* 13, e14–e25. doi: 10.62345/jads.2024.13.1.51
- Larrondo-Ureta, A., and Peña-Fernández, S. (2024). La formación de periodistas en la era de la inteligencia artificial: Aproximaciones desde la epistemología de la comunicación. *Ann. ThinkEP* 18:e18e1. doi: 10.3145/thinkep.2024.e18e1
- Larson, S., and Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Rev.* 9:1469. doi: 10.14763/2020.2.1469
- Leiser, M. R. (2022). "Bias, journalistic endeavours, and the risks of artificial intelligence" in Artificial intelligence and the media: Reconsidering rights and responsibilities. eds. T. Phibbs and A. Alén-Savikko (Cheltenham: Edward Elgar Publishing), 8–32.
- Lin, B., and Lewis, S. C. (2022). The one thing journalistic AI just might do for democracy. *Digit. Journal* 10, 1627–1649. doi: 10.1080/21670811.2022.2084131
- Lopezosa, C., Codina, L., Pont-Sorribes, C., and Villaz, M. (2023). Use of generative artificial intelligence in the training of journalists: challenges, uses and training proposal. *Prof. Inform.* 32:8. doi: 10.3145/pt.2023.jul.08
- Manfredi-Sánchez, J. L., and Ufario-Ruiz, M. J. (2020). Inteligencia artificial y periodismo: Una herramienta contra la desinformación. *Rev. CIDOB d'Àfers Int.* 124, 49–72. doi: 10.24241/rci.2020.124.149
- Micó, J. L., Casero-Ripollés, A., and García-Orosa, B. (2022). "Platforms in journalism 4.0: the impact of the fourth industrial revolution on the news industry" in *Total journalism. Studies in big data*. eds. J. Vázquez-Herrero, A. Silva-Rodríguez, M. C. Negreira-Rey, C. Touzal-Iñan and X. López-García, vol. 97 (Cham: Springer), 97–112.
- Milmo, D. (2023). ChatGPT reaches 100 million users two months after launch. *The Guardian*, 2 February. Available online at: <https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-openai-fastest-growing-app> (Accessed March 22, 2024).
- Moreno-Espinoza, P., Abdulrahman Alansyeh, R. A., and Figueroa Benítez, J. C. (2024). El Big Data y la inteligencia artificial como soluciones a la desinformación. *Deza Comunicación* 38, 437–451. doi: 10.31921/dezacom.38a2029
- Motlagh, N. Y., Khajavi, M., Sharifi, A., and Ahmadi, M. (2023). The impact of artificial intelligence on the evolution of digital education: a comparative study of OpenAI text generation tools including ChatGPT, Bing chat, Bard, and Ernie. *arXiv preprint*. Available online at: <https://arxiv.org/abs/2309.02029> (Accessed March 22, 2024).
- Noain Sánchez, A. (2022). Addressing the impact of artificial intelligence on journalism: the perception of experts, journalists and academics. *Commun. Soc.* 35, 105–121. doi: 10.15581/003.35.3.105-121
- Opdahl, A. L., Tøsse, B., Dang-Nguyen, D. T., Motta, E., Scity, V., Thorsen, E., et al. (2023). Trustworthy journalism through AI. *Data Knowl. Eng.* 146:102182. doi: 10.1016/j.datak.2023.102182
- Palacios Tapia, A. G. (2023). La influencia de la inteligencia artificial en el periodismo: Uso de herramientas y aplicaciones. Unpublished manuscript.
- Peña-Fernández, S., Mesa-Ayerdí, K., Larrondo-Ureta, A., and Díaz-Noci, J. (2023a). Without journalists, there is no journalism: the social dimension of generative artificial intelligence in the media. *Prof. Inform.* 32:e320227. doi: 10.3145/pt.2023.mar.27
- Peña-Fernández, S., Peña-Aleros, U., and Eizmendi-Iraola, M. (2023b). El discurso de los periodistas sobre el impacto de la inteligencia artificial generativa en la desinformación. *Estad. Mensaj. Period.* 29, 833–841. doi: 10.5209/comp.88673
- Pérez, F., Briones, B., Escobá, A., López, G., Maudos, J., and Pascual, F. (2023). Los medios de comunicación en la era digital. Fundación BBVA.
- Phelan, P. (2022). Are the current legal responses to artificial intelligence-facilitated 'deepfake' pornography sufficient to curtail the inflicted harm? *N. Engl. L. Rev.* 9, 20–45.
- Rubén, V. L. (2022). "Artificially Intelligent Solutions: Detection, Debunking, and Fact-Checking" in *Disinformation and Misinformation* (Cham: Springer). doi: 10.1007/978-3-030-95656-1_7
- Sánchez-García, P., Merayo-Álvarez, N., Calvo-Barbero, C., and Díez-Gracia, A. (2023). Spanish technological development of artificial intelligence applied to journalism: companies and tools for documentation, production and distribution of information. *Prof. Inform.* 32:e320208. doi: 10.3145/pt.2023.mar.08
- Singh, S. K., Kumar, S., and Mehra, P. S. (2023). ChatGPT and Google Bard AI: a review. In: *Proceedings of the 2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*. June, 1–6. IEEE.
- Schraewerd, S. J., Wu, Y. K., Hickerson, A., and Wright, M. (2024). "Dangers and deepfakes: using scenario-based role-play to study journalists' behaviour towards using AI-based verification tools for video content." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May, 1–17.
- Subela-Hernández, B. J., and Vizaño-Laorga, R. (2023). Retos y oportunidades en la lucha contra la desinformación y los derechos de autor en periodismo: MediaVerse (IA, blockchain y smart contracts). *Estad. Mensaj. Period.* 29:88081. doi: 10.5209/comp.88081
- Tejedor, S., and Vila, P. (2021). Easjournalism: a conceptual approach to a hybrid formula between journalism and artificial intelligence. *J. Med.* 2, 830–840. doi: 10.3390/journalmedia2040048
- Terol, T. M. (2023). Innovación mediática: Aplicaciones de la inteligencia artificial en el periodismo en España. *Tecnol. Vb. Media* 17, 41–60. doi: 10.56418/tat.17.1.2023.3
- Ufario-Ruiz, M. J., Calvo-Rubio, L. M., and Murcia Verdú, F. J. (2021). Los desafíos éticos del periodismo en la era de la inteligencia artificial. *Estudios sobre el Mensaje Periodístico* 27, 673–684. doi: 10.5209/comp.69708
- Ufario-Ruiz, M. J., Pietras-Geide, C., and Tüner, M. (2020). La enseñanza-aprendizaje del periodismo automatizado en instituciones públicas: Estudios, propuestas de viabilidad y perspectivas de impacto de la IA. *Análisis Qual. Comun. Cult.* 62, 131–146.
- Wu, S., Tandoc, E. C., and Salmon, C. T. (2019). Journalism reconfigured. *Journal. Stud.* 20, 1440–1457. doi: 10.1080/1461670X.2018.1521299

La inteligencia artificial como motor de innovación en los medios: apropiación técnica, cautela informativa

Barbara Sarrionandia

Universidad del País Vasco (UPV/EHU) 

Simón Peña-Fernández

Universidad del País Vasco (UPV/EHU) 

Jesús Ángel Pérez-Dasilva

Universidad del País Vasco (UPV/EHU) <https://dx.doi.org/10.5209/emp.102961>

Recibido: 25 de mayo de 2025 / Aceptado: 26 de junio de 2025

Resumen. La inteligencia artificial (IA) está transformando de forma profunda el trabajo en los medios de comunicación, alterando no solo las rutinas de producción, sino también los perfiles profesionales y la organización interna de las redacciones. Esta investigación examina cómo integran la IA los medios de comunicación del País Vasco, con el objetivo de conocer las prácticas innovadoras que desarrollan en este ámbito y comprender su impacto en las distintas fases del trabajo periodístico. Para ello, se ha adoptado una metodología mixta que combina entrevistas en profundidad a nueve responsables de seis medios vascos con una encuesta a 504 periodistas y profesionales en activo. El análisis se ha estructurado según las tres fases clásicas del proceso periodístico: búsqueda de información, producción de contenido y distribución. Los resultados muestran una adopción cauta de la IA en el plano informativo, pues su uso se orienta principalmente a tareas auxiliares, como la transcripción automática, la corrección de textos o la traducción. Aunque se han identificado iniciativas experimentales innovadoras por parte de los medios, como la clonación de voz en productos narrativos, estas se han desarrollado en contextos periféricos y no informativos. Asimismo, pese al incipiente desarrollo de estrategias corporativas, normativas internas y aplicaciones propias, la implementación de la inteligencia artificial continúa dependiendo en gran medida de las soluciones ofrecidas por las grandes plataformas tecnológicas y de la iniciativa individual de los profesionales.

Palabras clave. Inteligencia artificial, periodismo, innovación, medios de comunicación, transformación digital.

[ENG] Artificial Intelligence as a driver of media innovation: technical appropriation, editorial caution

Abstract. Artificial intelligence (AI) is profoundly transforming work in the media sector, reshaping not only production routines but also professional profiles and the internal organisation of newsrooms. This study examines how media outlets in the Basque Country are integrating AI, with the aim of identifying the innovative practices being developed and understanding their impact across the different stages of journalistic work. To this end, a mixed-methods approach was adopted, combining in-depth interviews with nine professionals from six Basque media organisations and a survey of 504 active journalists and media professionals. The analysis was structured around the three classic phases of the journalistic process: information gathering, content production, and distribution. The findings reveal a cautious adoption of AI in news-related activities, as its use is primarily focused on auxiliary tasks such as automatic transcription, text editing, or translation. While some innovative experimental initiatives have been identified –such as voice cloning in narrative products– these have mostly emerged in peripheral, non-news contexts. Moreover, despite the early development of corporate strategies, internal regulations, and bespoke applications, the implementation of artificial intelligence largely continues to rely on solutions provided by major tech platforms and on the individual initiative of professionals.

Keywords. Artificial intelligence, journalism, innovation, media, digital transformation.

Cómo citar: Sarrionandia, B., Peña-Fernández, S., y Pérez-Dasilva, J. A. (2025). La inteligencia artificial como motor de innovación en los medios: apropiación técnica, cautela informativa. *Estudios sobre el Mensaje Periodístico*, 31(3), 603-613. <https://dx.doi.org/10.5209/emp.102961>

1. Introducción

La integración de la inteligencia artificial (IA) en los medios de comunicación, lejos de ser una simple innovación técnica, afecta de manera transversal a la producción, distribución y recepción de los contenidos, por lo que reconfigura tanto las rutinas profesionales como los valores editoriales (Peña-Fernández *et al.*, 2023). Desde la irrupción y generalización del uso de las aplicaciones de la IA generativa, esta tecnología ha pasado de un lugar marginal en los procesos editoriales a convertirse en un eje que reconfigura las rutinas de trabajo, los marcos organizativos y las competencias profesionales del periodismo contemporáneo (Chan-Olmsted, 2019; Liu, 2024).

Lejos de implantarse de forma homogénea o lineal, la incorporación de la IA presenta formas, ritmos y resistencias diferenciadas según el tipo de medio, su modelo de financiación o su grado de digitalización previo. Al igual que ha ocurrido en otras fases de la digitalización mediática (Boczkowski, 2004), su adopción no puede desvincularse de factores estructurales, como la disponibilidad de recursos, la cultura organizacional o el posicionamiento estratégico de cada medio (Al-Jenaibi y Bulhoon, 2024).

Para conocer mejor esta realidad, han comenzado a desarrollarse marcos que permiten sistematizar el análisis y la comparación de los usos de la IA en los entornos periodísticos, con el objetivo de evaluar el grado de innovación, la orientación estratégica y la madurez institucional en su implementación (Gondwe, 2023).

En este contexto, la presente investigación propone un análisis de la incorporación de la IA en medios de comunicación del País Vasco. Para ello, se plantean las siguientes preguntas de investigación:

PI1. ¿Cuál es el uso de la IA por parte de los periodistas vascos? ¿Qué aplicaciones utilizan y para qué fines?

PI2. ¿Cuál es el uso diferencial que se realiza de estas aplicaciones en función del tipo de medio de comunicación y de sus perfiles profesionales?

PI3. ¿Cuál es el uso de la IA en relación a las fases clásicas del proceso periodístico (búsqueda de información, producción de contenido y distribución o relación con la audiencia)?

PI4. ¿Qué tipo iniciativas innovadoras y aplicaciones propias han desarrollado los medios de comunicación vascos en este ámbito?

2. Marco teórico

Una de las transformaciones más relevantes impulsadas por la IA en los medios de comunicación ha sido la progresiva automatización de tareas tradicionalmente desempeñadas por periodistas y profesionales. Así, herramientas para la redacción automática de noticias, la generación de titulares, el análisis masivo de datos o la personalización algorítmica de contenidos han comenzado a redefinir el flujo editorial de las redacciones (Noain-Sánchez, 2022). De hecho, según el estudio de Sonni *et al.* (2024), tres de cada cuatro medios ya ha implementado alguna forma de redacción automatizada, lo que indica no solo una adopción amplia, sino también una tendencia hacia la naturalización de estas herramientas.

Con la transformación de los procesos de producción y difusión de noticias, la inteligencia artificial se

presenta como una herramienta con un gran potencial para mejorar la eficiencia e impulsar la innovación en los medios de comunicación (Bakke y Barland, 2022; Danzon-Chambaud, 2021; Orosa *et al.*, 2023; Terol, 2023), ya sea mediante la introducción de nuevas tecnologías y rutinas productivas, como a través de la generación de formatos narrativos y modelos de negocio disruptivos (Bleyen *et al.*, 2014; Westlund y Lewis, 2014). Esta perspectiva permite considerar la IA no solo como una herramienta técnica, sino como un motor de innovación que modifica la cultura profesional y la relación entre periodistas, audiencias y plataformas digitales (García Avilés, 2021; Westlund y Lewis, 2014).

Sin embargo, su uso también conlleva la aparición de nuevas amenazas y fragilidades en el ecosistema informativo (Brennen *et al.*, 2018), por lo que es necesario generar estructuras internas de gobernanza en los medios que permitan articular normas propias de uso responsable (Al-Zoubi *et al.*, 2024; Bayer, 2024; Benjamins, 2020; Cath, 2018; de-Lima-Santos *et al.*, 2024; Lewis y Moorkens, 2020).

La incorporación de IA también está transformando los perfiles profesionales y ha propiciado el surgimiento de figuras profesionales híbridas como el curador algorítmico o el periodista de datos, cuya formación implica tanto saberes comunicativos como habilidades técnicas avanzadas (Lopezosa *et al.*, 2023; Tüñez-López *et al.*, 2021). En un entorno de naturaleza crecientemente híbrida, este cambio demanda inversiones institucionales significativas en formación y desarrollo profesional, así como políticas internas de transición laboral que acompañen estos procesos (Sarrionandia *et al.*, 2025; Verma, 2024).

A mayor escala, la progresiva integración de la IA está propiciando una reconfiguración organizativa visible, que se manifiesta en la creación de unidades internas especializadas, principalmente en los grandes medios. Estas unidades actúan como nodos de gobernanza tecnológica, encargados de identificar, evaluar y regular el uso de herramientas de IA, al tiempo que contribuyen a definir criterios éticos y editoriales para su implementación (Feiras-Ceide *et al.*, 2022; Schiff *et al.*, 2024).

En este contexto, la incorporación integral de la IA se está desarrollando, desde el punto de vista de los procesos informativos, fundamentalmente en las fases de búsqueda y distribución de noticias (Sánchez García *et al.*, 2023). La optimización de los procesos existentes se ha centrado hasta el momento en actividades técnicas o complementarias, pues los profesionales insisten en mantener el control en todas las etapas, particularmente en la selección y edición de contenidos, lo que refleja su voluntad de preservar su rol como garantes del proceso informativo (Wu *et al.*, 2018).

Por el contrario, los responsables de los medios promueven más activamente su integración en todas las fases del trabajo de los medios, como una vía de aliviar la precaria situación económica de muchas de sus empresas (Kim y Kim, 2017). Sin embargo, los altos costes de desarrollo de aplicaciones propias suponen una barrera para su implantación, que provoca que en su mayor parte la IA suponga una dependencia aún mayor de los medios de comunicación respecto de las grandes plataformas tecnológicas (Broussard *et al.*, 2019; Simon, 2022).

3. Metodología

3.1. Marco metodológico

Para el análisis de las aplicaciones de la IA en los medios de comunicación, una de las taxonomías más ampliamente difundidas las clasifica según las fases del proceso periodístico: búsqueda de información, producción de contenido y distribución o relación con la audiencia (Gutiérrez-Caneda *et al.*, 2023; loscote *et al.*, 2024; Sánchez-García *et al.*, 2023). Esta sistematización no solo permite identificar en qué momentos del proceso editorial se están aplicando las herramientas de IA, sino que también posibilita conocer qué medios son más proclives a usos experimentales o estructurales y qué otros optan por una integración más instrumental, limitada a tareas técnicas o de soporte.

El estudio de la IA en los medios también se puede basar en los fundamentos computacionales de las herramientas aplicadas: aprendizaje automático, procesamiento del lenguaje natural, visión por computador, reconocimiento de voz, planificación/optimización, sistemas expertos y robótica (de-Lima-Santos y Ceron, 2021). Esta clasificación permite conectar las tecnologías concretas con las tareas que desempeñan dentro del proceso editorial, con lo que revela niveles de especialización, dependencia tecnológica y requerimientos de infraestructura.

Un tercer enfoque, de naturaleza más conceptual, organiza las herramientas de IA según la función o finalidad que cumplen (Gondwe, 2023). Así, el uso de la IA podría analizarse en cuatro dimensiones: generación de contenido, análisis de datos, automatización de tareas y participación de la audiencia. Con ella puede distinguirse entre usos instrumentales (por ejemplo, etiquetado y transcripción) y usos creativos o simbólicos (como la clonación de voz o la personalización narrativa).

Noain-Sánchez (2022), por su parte, diferencia entre periodismo automatizado, algorítmico y aumentado en función del grado de intervención humana y del rol que juegan los sistemas computacionales en la producción informativa. Esta distinción resulta particularmente útil para interpretar las tensiones entre

eficiencia, control editorial y credibilidad en entornos donde conviven la producción manual, la automatización parcial y la inteligencia generativa (Epstein y Hertzmann, 2023; Franganillo, 2023).

A partir de estas taxonomías, el presente estudio adopta como marco principal de análisis la clasificación funcional por fases del proceso periodístico: búsqueda de información, producción de contenido y distribución. Esta estructura permite ubicar cada herramienta tecnológica dentro de una secuencia editorial reconocible, por lo que facilita el análisis transversal entre medios, tecnologías y funciones.

3.2. Diseño del estudio

Esta investigación adopta un enfoque mixto que combina técnicas cualitativas y cuantitativas con el objetivo de comprender los usos de la IA en los medios de comunicación del País Vasco. La integración de ambos enfoques metodológicos responde a la necesidad de conocer, por un lado, las estrategias desarrolladas por los medios en relación con la IA, y por otro, de evaluar la actitud de los profesionales del periodismo ante estas tecnologías emergentes (Peña-Fernández *et al.*, 2023).

En el ámbito cualitativo, se han realizado de nueve entrevistas en profundidad semiestructuradas a profesionales de seis medios de comunicación con implantación en el País Vasco. Los perfiles seleccionados (Tabla 1) ocupan cargos de responsabilidad en áreas relacionadas con innovación tecnológica, la inteligencia artificial, la dirección editorial o el desarrollo de producto, lo que ha permitido conocer las aplicaciones que se están desarrollando en este ámbito y la visión estratégica que las impulsa.

Las entrevistas se realizaron entre enero y abril de 2025, tuvieron una duración media de una hora y fueron grabadas, transcritas y analizadas mediante codificación temática. El análisis de contenido se llevó a cabo tomando como base las tres fases clásicas del flujo de trabajo periodístico y se realizó mediante codificación temática abierta y axial. Asimismo, se alcanzó la saturación teórica al constatar que en las últimas entrevistas no emergían nuevas categorías relevantes.

Tabla 1. Entrevistas en profundidad.

Código	Género	Medio	Tipo	Cargo
I1	H	Vocento / El Correo	Prensa	Responsable de nuevas tendencias tecnológicas
I2	H	EITB	Radiotelevisión pública	Director de Data, IA e Innovación
I3	H	Grupo Noticias / Dela	Prensa	Dirección del área tecnológica
I4	H	Naiz	Prensa	Director adjunto
I5	H	Berria	Prensa	Responsable de tecnologías de la información
I6	H	Cadena SER / Radio Bilbao	Radio	Redactor
I7	M	EITB	Radiotelevisión pública	Subdirectora de Labs - Entretenimiento
I8	M	Artefacto	Productora	Cofundadora
I9	M	Prisa Media	Radio, online, prensa	Coordinadora de Voz

Fuente: elaboración propia.

El apartado cuantitativo de la investigación se basa en una encuesta *online* con apoyo telefónico realizada durante los meses de mayo y junio de 2024, dirigida a una muestra de 504 periodistas y trabajadores de medios de comunicación en activo del País Vasco. La muestra se construyó a partir de dos fuentes principales: la Guía de Comunicación Abierta del Gobierno Vasco (<https://gida.irekia.euskadi.eus>), que proporciona una base de datos completa de profesionales y medios que operan en la comunidad, y el listado de miembros de la Asociación Vasca de Periodistas. El cuestionario fue diseñado para garantizar el anonimato y la confidencialidad de las respuestas.

En términos demográficos, la muestra está compuesta por 276 hombres (55,1 %), 223 mujeres (44,5 %) y dos personas que se identificaron en la categoría de «otros» (0,4 %). Por tipos de medios, 199 encuestados trabajaban en prensa (39,5 %), 83 en radio (16,5 %), 69 en televisión (13,7 %), 66 en medios digitales (13,1 %) y 87 (17,3 %) en gabinetes de comunicación. En lo que se refiere a las actividades que realizaban dentro del medio, 313 (62,1 %) trabajaban como redactores, 95 en tareas de locución o presentación (18,8 %), 38 en el área técnica (7,5 %) y 58 en el área corporativa (11,5 %). Esta diversidad permitió obtener una visión representativa del ecosistema mediático vasco y evaluar posibles patrones diferenciales de percepción según género, medio, rol o antigüedad.

4. Resultados

4.1. Aplicaciones y usos de la IA

La encuesta realizada permite determinar, en primer lugar, que el uso aplicaciones de IA está bastante extendido entre los periodistas y los profesionales de los medios de comunicación, pues el 56,5 % de los encuestados declara utilizar algún tipo de herramienta basada en la IA para su trabajo. No hay, curiosamente, diferencia alguna en función de la trayectoria profesional o la edad, pues los periodistas con más de 20 años de experiencia utilizan estas aplicaciones casi en idéntico porcentaje que los jóvenes profesionales

con menos de 5 años en el trabajo (55,2 % frente a 57,7 %, respectivamente). Estos datos indican que, al menos en los medios de comunicación, el uso de la IA no está capitalizado por profesionales jóvenes o con un perfil nativo digital, sino que ha sido adoptado igualmente por profesionales con larga trayectoria, lo que refuerza la idea de una apropiación transversal de esta tecnología. Una diferencia algo mayor puede apreciarse en términos de género, pues las mujeres utilizan la IA en mayor medida que los hombres (62,4 % frente a 51,4 %).

Si nos fijamos en las herramientas más populares (Tabla 2), ChatGPT lidera con claridad entre las aplicaciones empleadas por parte de periodistas y profesionales, y dos de cada cinco de ellos (40,1 %) lo utilizan para su trabajo, seguido por Canva (33,5 %) —herramienta de diseño gráfico *online* que, mediante inteligencia artificial, permite crear contenido visual— y Gemini (8,8 %). A ellas se une una larga lista de aplicaciones —entre las que, además de las ya mencionadas, se citan Copilot, DeepL, HappyScribe, TRINT, TurboScribe, etc.— que dibujan un panorama de una enorme diversidad.

La apropiación de estas herramientas en función del tipo de medio para el que se trabaja y el perfil del puesto de trabajo también arroja datos reseñables. Así, los profesionales que se dedican a tareas técnicas o relacionadas con la comunicación corporativa utilizan la IA —por término medio— el doble que los perfiles más asociados al periodismo (redacción, presentación, locución, etc.). El uso de la IA está ya, por tanto, muy presente en las tareas periféricas y de soporte a la actividad periodística, que sin embargo se muestra más renuente a utilizar de forma generalizada estas aplicaciones. Por tipos de soportes, los resultados son muy similares, pues los profesionales que trabajan en los medios nativos digitales, las redes sociales y los gabinetes de comunicación emplean en mucho mayor medida estas herramientas que los trabajadores de los soportes periodísticos tradicionales (prensa, radio, televisión).

Tabla 2. Aplicaciones IA utilizadas por periodistas y profesionales de los medios de comunicación vascos, por tipo de medios.

Aplicaciones	Total	%	Prensa	Radio	TV	Digital	Gabin.	Redac.	Locu.	Tec.	Corp.
ChatGPT	201	40,1 %	28,6 %	33,7 %	29,0 %	59,1 %	66,7 %	37,4 %	26,3 %	60,5 %	64,7 %
Canva	168	33,5 %	27,1 %	25,3 %	21,7 %	56,1 %	47,1 %	32,9 %	24,2 %	47,4 %	41,2 %
Gemini	44	8,8 %	6,5 %	12,0 %	13,0 %	9,1 %	8,0 %	8,0 %	8,5 %	10,5 %	13,7 %
Dall-E	26	5,2 %	4,0 %	3,6 %	2,9 %	13,6 %	5,2 %	5,8 %	1,1 %	13,2 %	3,9 %
Midjourney	13	2,6 %	2,5 %	1,2 %	1,4 %	3,3 %	5,7 %	1,9 %	0,0 %	7,9 %	7,8 %
Otras	74	14,8 %	17,1 %	6,0 %	8,7 %	17,7 %	20,7 %	16,0 %	6,3 %	13,2 %	21,6 %
Ninguna	219	43,5 %	49,2 %	59,0 %	58,0 %	19,7 %	21,8 %	43,1 %	61,1 %	31,6 %	21,6 %

Fuente: elaboración propia.

Las tareas para las que se utilizan las herramientas de la IA refuerzan este perfil accesorio y complementario a la actividad periodística (Tabla 3). Así, los usos más habituales son la traducción automática de textos (46,4 %), las consultas temáticas (preguntas, datos, etc.) (30 %) o la corrección y revisión de textos (ortografía, estilo, etc.) (27,6 %). Por el contrario, los usos explícitamente generativos son mucho menos habituales, en particular entre los perfiles más periodísticos. La generación de textos alcanza

el 21 % globalmente, pero apenas sobrepasa el 13 % entre los profesionales de los soportes tradicionales, una cifra que se triplica en el caso de los medios nativos digitales, las redes sociales y los gabinetes de comunicación.

Si en lugar de los soportes nos fijamos en las tareas, ocurre algo muy similar, pues los trabajadores en el ámbito de la redacción y de la locución/presentación recurren a la generación de textos en mucha menor medida que el personal técnico y los trabaja-

dores del área corporativa. En conjunto, también se recurre en menor medida al uso de la IA para la generación de imágenes (16,5 %) y sonidos (3,0 %). Por tanto, el uso aún limitado de la IA en la generación directa de contenido sugiere una actitud prudente en su adopción, donde prevalece el criterio profesional y la autoría humana, especialmente en contextos informativos.

En otras palabras, los datos demuestran que la adopción y el uso de la IA en los medios de comunicación es mayor cuanto más alejado está el trabajo que se realiza de las labores específicamente perio-

dísticas, mientras que su utilización se multiplica en las tareas periféricas o en la información no específicamente periodística, como la que elaboran los gabinetes de comunicación.

Tampoco sorprende que entre los adoptadores tempranos de estas aplicaciones destaquen los trabajadores de los medios nativos digitales y las redes sociales, en cuya rápida apropiación de esta tecnología puede influir su perfil más próximo a las cuestiones técnicas o el trabajo en estructuras empresariales más pequeñas, que incentivan la optimización de los recursos.

Tabla 3. Usos de la IA y usos realizados por periodistas y profesionales de los medios de comunicación vascos.

Aplicaciones	Total	%	Prensa	Radio	TV	Digital	Gabin.	Redac.	Locu.	Tec.	Corp.
Traduc. autom.	234	46,4 %	43,7 %	37,3 %	33,3 %	63,6 %	58,6 %	44,4 %	33,7 %	65,8 %	64,7 %
Consult. temát.	151	30,0 %	20,6 %	26,5 %	23,2 %	43,9 %	40,4 %	25,9 %	23,2 %	42,1 %	58,8 %
Correc. textos	139	27,6 %	29,1 %	25,3 %	15,9 %	24,2 %	37,9 %	26,2 %	22,1 %	34,2 %	37,3 %
Transcripción	116	23,0 %	29,1 %	18,1 %	8,7 %	22,7 %	25,3 %	24,9 %	13,7 %	26,3 %	23,5 %
Gener. textos	106	21,0 %	13,6 %	13,3 %	13,0 %	42,4 %	35,6 %	19,2 %	9,5 %	44,7 %	35,3 %
Gener. imágenes	83	16,5 %	13,1 %	14,5 %	17,4 %	25,8 %	18,4 %	15,3 %	12,6 %	26,3 %	21,6 %
Subtitulado	72	14,3 %	8,5 %	8,4 %	20,3 %	16,7 %	26,4 %	11,5 %	10,5 %	34,2 %	19,6 %
Anál. datos	72	14,3 %	11,1 %	8,4 %	8,7 %	25,8 %	23,0 %	14,4 %	8,4 %	13,2 %	23,5 %
Verif. datos	48	9,5 %	6,5 %	7,2 %	8,7 %	18,2 %	12,6 %	8,6 %	6,3 %	15,8 %	13,7 %
Gener. audio	15	3,0 %	2,0 %	6,0 %	1,4 %	1,5 %	4,6 %	1,9 %	4,2 %	2,6 %	7,8 %
Otras	12	2,4 %	3,0 %	2,4 %	3,9 %	0,0 %	2,3 %	2,2 %	1,1 %	5,3 %	2,0 %
Ninguna	219	43,5 %	49,2 %	59,0 %	58,0 %	19,7 %	21,8 %	43,1 %	61,1 %	31,6 %	21,6 %

Fuente: elaboración propia.

4.2. Usos de la IA en la búsqueda de información

Si nos fijamos en el uso de la IA en función de las fases clásicas del proceso periodístico, observamos que la traducción automática, la búsqueda de información y la corrección de textos son los usos dominantes. Se puede así acceder a fuentes multilingües y filtrar grandes volúmenes de datos con mayor rapidez, por lo que se agilizan tareas rutinarias y repetitivas.

Al margen de estos usos profesionales en el trabajo diario, también se han identificado a través de las entrevistas en profundidad iniciativas innovadoras a nivel empresarial para el desarrollo de herramientas propias de IA. Así, para la primera fase del proceso informativo —la búsqueda de información— el grupo EITB ha desarrollado una estrategia centralizada para evaluar las tecnologías emergentes, incluida esta primera fase del trabajo informativo. Aunque aún no se han integrado de forma generalizada en las rutinas de las redacciones, se ha iniciado la exploración de herramientas de minería de datos y de búsqueda documental automatizada para los medios de comunicación del grupo (12). La cautela con la que se está desarrollando este proceso responde al principio de supervisión humana y trazabilidad, recogido en el decálogo institucional publicado por EITB en enero de 2024, que establece que toda herramienta empleada debe ser previamente testada y validada por el grupo técnico interdisciplinar.

El acceso documental y la recolección de información es también la función principal de la herramienta de IA que está desarrollando el diario *Berria*, «un asistente en forma de chat que ayude en la búsqueda de contenidos propios» (15). También en este caso, esta herramienta representa una apuesta por avanzar hacia formas de recuperación automatizada de información que faciliten la localización eficiente de materiales internos en grandes volúmenes de datos.

4.3. Producción: automatización y experimentación

4.3. Producción: automatización y experimentación

El uso de la IA en fase de producción, entendida como el momento central en la elaboración de contenidos informativos y no informativos, evidencia un desarrollo aún incipiente, pero con implicaciones significativas para los medios de comunicación. A partir de las entrevistas realizadas a responsables tecnológicos, jefaturas editoriales y periodistas especializados, han podido identificarse dos tendencias predominantes: por un lado, la incorporación paulatina de herramientas de apoyo a la generación de contenido multimedia y, por otro, el desarrollo de experimentaciones narrativas orientadas a explorar las capacidades expresivas de la IA, con supervisión humana.

Un ejemplo de estas aproximaciones experimentales al uso de la IA es el podcast *Ahalaken*, producido por EITB junto con las productoras Trebe y Artefacto. En este proyecto, la fase de producción estuvo marcada por dos componentes innovadores: la clonación de la voz del meteorólogo Peio Zabala —fallecido en 2022— y la generación automática de videos a partir del contenido sonoro. La voz fue recreada mediante una herramienta neuronal de síntesis de voz a partir de grabaciones del archivo

radiofónico de la casa, con una herramienta TTS entrenada específicamente para mantener la entonación, la dicción y el acento del locutor original. El resultado fue una narración ficcional con frases limitadas, incorporadas en una pieza de ficción sonora íntegramente en euskera. El proyecto se llevó a cabo con autorización explícita de la familia de Zabala y de la orden de Arantzazu —era fraile franciscano—, aplicando un protocolo ético estricto (12, 17).

En paralelo, para la generación automatizada de vídeo en Ahalaken se combinaron múltiples herramientas de IA generativa visual —entre ellas Runway, Klingo y Halo— especializadas en *lip-sync*, animación facial y movimiento corporal. El objetivo era alcanzar un grado de realismo y expresividad que permitiera «mimetizar los gestos humanos con la mayor fidelidad posible» (18). Para ello, se optó por un enfoque de planos cortos, evitando planos generales donde los modelos generativos presentan aún problemas de coherencia visual. La complejidad técnica de este proceso exigió la intervención de un equipo multidisciplinar y una labor de postproducción intensiva para homogeneizar color, contraste y textura.

Radio Bilbao también ha desarrollado una iniciativa similar en la recuperación narrativa y la memoria cultural. En 2023, y con motivo de su 90 aniversario, el equipo de SER Euskadi elaboró una pieza sonora —concretamente una entrevista— utilizando una voz sintética basada en archivos del músico vasco Kepa Junkera, incapacitado para hablar desde 2018. El proceso incluyó la limpieza y alineamiento fonético de entrevistas antiguas para entrenar un modelo TTS de síntesis vocal con capacidad de reproducir la prosodia de Junkera, en la que la voz generada digitalmente iba leyendo las respuestas dadas por el músico en el propio estudio. La producción contó con el consentimiento explícito del propio artista y fue presentada al público con un aviso claro sobre el uso de tecnologías de voz sintética. «El objetivo era rendir homenaje, no suplantar su discurso» (19), lo que subraya el carácter limitado y simbólico de la acción. La pieza no fue difundida como contenido informativo ni se integró en las rutinas productivas ordinarias, sino que fue concebida como un ejercicio narrativo singular, enmarcado en las líneas de innovación del grupo.

4.4. Edición: transcripción, traducción y corrección automatizadas

La fase de edición constituye una de las áreas donde la incorporación de herramientas de IA está comenzando a adquirir una relevancia creciente en las rutinas de los medios de comunicación vascos. Entre las tareas más afectadas por la automatización se incluyen la transcripción, la traducción, el subtítulo, el resumen automático de textos, la corrección ortotipográfica y el tratamiento de archivos multimedia.

En los medios analizados, una de las aplicaciones más extendidas es la transcripción automática de archivos sonoros. En el caso de EITB, la transcripción se encuentra integrada en la plataforma de audio Guau, que permite ver en tiempo real la transcripción de los audios en el subtítulo, lo que representa un avance significativo en términos de accesibilidad.

En su desarrollo, el proyecto no ha carecido de las limitaciones técnicas asociadas al reconocimiento del euskera: «Cuando tienes una entrevista, unas intervenciones, en función del dialecto que utilice complica también el uso» (12).

En el caso del diario *Berría*, la transcripción automática ha sido objeto de ensayos previos, interrumpidos temporalmente por problemas de calidad, pero reactivados recientemente ante mejoras sustanciales detectadas en los servicios utilizados (15). Como en otros casos, la postura sobre la integración de la IA ha sido prudente en este caso, pero abierta hacia la integración de herramientas que permitan reducir la carga de trabajo manual y optimizar los tiempos de edición.

En el caso del diario *Naiz*, el uso de IA se orienta principalmente a funciones de apoyo editorial, sin que la generación autónoma de contenidos forme parte de sus rutinas productivas. Este diario utiliza herramientas como Marfeel, basada en Copilot, para el análisis de audiencias y la mejora de titulares, así como extensiones para la corrección automática y la transcripción de audio, desarrolladas en colaboración con Elhuyar (Trebe). Asimismo, se ha implementado iVoice Scans como sistema de verificación de autenticidad de archivos sonoros (14). Aunque se están desarrollando proyectos más ambiciosos —como un traductor neuronal para publicar en euskera o un sintetizador de voz para lectura automatizada de textos—, la política editorial del medio prohíbe el uso de IA generativa en los productos informativos finales, y todas las traducciones o transformaciones automáticas deben ser verificadas por periodistas antes de su publicación.

Otro de los usos identificados de la IA en los medios de comunicación vascos en la fase de edición ha sido la traducción automática. En EITB, estas funcionalidades se están probando de manera experimental en un servicio en la nube que incluye traducción, subtítulo, segmentación y metadatos automático de vídeos y audios, que busca facilitar la gestión de archivos audiovisuales multilingües y agilizar su edición para distintos canales y públicos (12).

Berría, por su parte, ha incorporado esta línea en un proyecto en colaboración con otros medios en euskera, como Hamarka Telebista, Euskalerria Irratia y Euskal Irratiak, y el apoyo de Elhuyar. El objetivo es desarrollar herramientas propias de traducción automática entre el euskera y otras lenguas que permitan alcanzar un alto grado de precisión, respetando las particularidades sintácticas, léxicas y morfológicas del idioma, y que contribuyan a garantizar una representación adecuada del euskera en el ecosistema digital emergente (15). Estas herramientas forman parte de estrategias más amplias vinculadas a la accesibilidad lingüística, aunque los responsables reconocen que persisten retos técnicos derivados de la complejidad morfosintáctica del idioma y de la escasez de corpus amplios para su entrenamiento. Por ello, el uso de correctores automáticos sigue dependiendo en gran medida de la supervisión humana, especialmente en entornos con altos estándares de calidad y control lingüístico como los medios de comunicación.

Aunque todavía de forma incipiente, algunas iniciativas de los medios de comunicación vascos apuntan a la posibilidad de aplicar modelos de IA

para generar resúmenes automáticos de textos periodísticos. Este tipo de aplicación ha sido mencionado como objetivo de desarrollo futuro dentro del consorcio liderado por Berria y apoyado por fondos europeos. La idea es incorporar resúmenes automáticos en procesos de indexación documental y distribución de contenidos, lo que permitiría ofrecer versiones abreviadas de artículos extensos para consumo rápido en plataformas digitales o redes sociales (I5). Si bien esta funcionalidad aún no ha sido implementada de manera sistemática en los medios analizados, su presencia en la hoja de ruta tecnológica de algunos de ellos indica un interés creciente por automatizar parcialmente tareas de redacción, sin comprometer la autoría ni la integridad del contenido.

5. Discusión y conclusiones

El impacto de la IA no puede entenderse al margen de las tensiones que atraviesan a los medios en la actualidad. Estas tecnologías no son neutrales ni universales en su implementación, sino que están moldeadas por intereses institucionales, contextos culturales específicos y capacidades técnicas desiguales, lo que da lugar a procesos de adopción heterogéneos y sujetos a múltiples condicionantes (Beckett et al., 2023). Su apropiación por parte de los medios obedece, por tanto, a múltiples factores como la disponibilidad de recursos, la cultura organizativa o, incluso, su idioma vehicular. En el caso de los medios regionales y locales, además, sus estructuras organizativas más reducidas y su menor capacidad de inversión tecnológica influyen directamente en la forma en que se incorporan tecnologías emergentes (Olsen y Hess, 2024).

Nos encontramos, por tanto, ante una transformación estructural cuyo desarrollo es lento y heterogéneo, en un contexto en el que la actitud de los medios de comunicación frente a la transformación digital ha sido, tradicionalmente, reactiva, defensiva y pragmática (Boczkowski, 2004). Ante la irrupción de innovaciones tecnológicas, los medios han tendido a replicar las prácticas de sus competidores, a protegerse frente a las iniciativas de las grandes compañías tecnológicas y a centrarse en amenazas inmediatas más que en oportunidades estratégicas a largo plazo.

El análisis de la implantación de la IA en los medios de comunicación vascos y de las iniciativas para desarrollar aplicaciones propias muestran, en primer lugar, una adopción prudente, selectiva y subordinada al criterio profesional. Si bien existen experiencias experimentales innovadoras como el podcast *Ahalaken* de EITB o el proyecto con voz sintética de la entrevista a Kepa Junkera de Cadena SER Bilbao, en ambos casos la IA se ha utilizado en espacios acotados y no informativos, con autorización expresa de las personas implicadas y advertencias explícitas sobre su naturaleza artificial (Barnett, 2023; Gupta et al., 2024). Se trata, por tanto, de iniciativas de naturaleza periférica y experimental, que se sitúan lejos del epicentro de la actividad cotidiana de estos medios.

Asimismo, el uso de las herramientas por parte de los profesionales de los medios de comunicación vascos refleja que, si bien su adopción es ya signifi-

cativa en ciertas tareas —especialmente en la fase de edición—, su uso generativo es mucho más limitado. El uso de la IA se concentra en tareas técnicas y auxiliares, donde actúa como tecnología de apoyo que mejora la productividad y calidad del trabajo, y no se traslada al ámbito informativo como un sustituto de la labor periodística. Esta diferencia en la adopción de la IA puede estar relacionada, en parte, con el escepticismo que persiste entre los periodistas sobre sus implicaciones para la profesión (Peña-Fernández et al., 2023).

Hasta el momento, los medios vascos operan con una fuerte dependencia de herramientas de terceros —con el polivalente ChatGPT como herramienta más utilizada—, aunque empiezan a surgir desarrollos propios o semipropios, especialmente en los grupos mediáticos de mayor tamaño o a través de consorcios. Sin embargo, al margen de estas iniciativas puntuales de naturaleza aproximativa o experimental, el uso cotidiano de la IA en las redacciones está mediatizado por las aplicaciones de las grandes plataformas comerciales (Nielsen y Ganter, 2022; Simon, 2022).

También se ha constatado una correlación entre la adopción de IA y la existencia de estructuras internas de gobernanza tecnológica. Medios como EITB han formalizado decálogos de uso, comisiones interdepartamentales y circuitos de autorización ética, mientras que otros, como Berria o Vocento, avanzan en la creación de documentos normativos o grupos de trabajo. Esto indica que la IA no se adopta de forma exclusivamente técnica, sino también de forma institucional y estratégica, por lo que aquellos medios que carecen de estructuras formales en este ámbito tienden a implementarla de forma mucho más tímida, sin desarrollos propios.

Los discursos de los entrevistados y los patrones de uso identificados permiten concluir también que la apropiación de la IA ha estado hasta el momento más asociada a la iniciativa personal de los trabajadores que a procesos de capacitación institucional. Salvo en el caso de EITB, que ha impulsado formaciones transversales, la mayoría de los medios no ha desplegado programas sistemáticos de alfabetización en IA, por lo que la formación ha corrido, en la mayor parte de los casos, de la mano de los propios profesionales (Sarrionandia et al., 2025).

Por último, la perspectiva de futuro que emerge de los discursos recogidos sugiere una voluntad de avanzar en la incorporación progresiva de la IA, especialmente en tareas de edición y distribución, pero bajo la premisa de que se mantenga la supervisión humana en la generación de contenidos y el control editorial. Los casos de EITB y SER, aunque pioneros, se conciben como experimentos dirigidos y no como modelos replicables sin adaptación.

En definitiva, se observa en los medios de comunicación una integración funcional de la IA, principalmente en tareas auxiliares y como herramienta de apoyo, así como como herramienta para la creación de contenidos experimentales, promocionales o editorialmente poco sensibles, mientras los contenidos informativos y periodísticos se mantienen en mayor medida aislados de su uso. Su apropiación temprana por parte de los medios digitales y los gabinetes de comunicación, por el contrario, sugiere que la optimización de los recursos, en particular en

las tareas más ajenas a las consideradas específicamente periodísticas, puede actuar como acelerador para su adopción.

Los hallazgos de este estudio subrayan la necesidad de que medios y profesionales consoliden políticas internas que regulen el uso responsable de la inteligencia artificial y que aseguren en todo momento la supervisión humana en la generación de contenidos. La apuesta por la formación continua en competencias digitales y el desarrollo de soluciones propias que reduzcan la dependencia de plataformas externas se perfilan como pautas clave para la implementación de la IA generativa en el ámbito mediático. Con ello, podrán optimizarse procesos

técnicos sin comprometer la calidad informativa ni la autonomía editorial.

6. Financiación y apoyos

Este artículo forma parte de la producción científica de los proyectos «Impacto de la inteligencia artificial y los algoritmos en los cibermedios, los profesionales y las audiencias», financiado por el Ministerio de Ciencia e Innovación (PID2022-1383910B-I00) y «Impacto de la inteligencia artificial en los medios de comunicación vascos y sus profesionales» (US 23/10), financiado por la UPV/EHU. También forma parte de la producción científica del grupo de investigación consolidado del Gobierno Vasco Gureiker (IT1496-22).

7. Contribución de autores

Conceptualización	Ideas; formulación o evolución de los objetivos y metas generales de la investigación.	Autores 1, 2 y 3
Curación de datos	Actividades de gestión para anotar (producir metadatos), depurar datos y mantener los datos de la investigación (incluido el código de software, cuando sea necesario para interpretar los propios datos) para su uso inicial y su posterior reutilización.	Autores 1, 2 y 3
Análisis formal	Aplicación de técnicas estadísticas, matemáticas, computacionales u otras técnicas formales para analizar o sintetizar datos de estudio.	Autores 1, 2 y 3
Adquisición de fondos	Adquisición del apoyo financiero para el proyecto que conduce a esta publicación.	Autores 2 y 3
Investigación	Realización de una investigación y proceso de investigación, realizando específicamente los experimentos, o la recolección de datos/evidencia.	Autores 1, 2 y 3
Metodología	Desarrollo o diseño de la metodología; creación de modelos.	Autores 1, 2 y 3
Administración del proyecto	Responsabilidad de gestión y coordinación de la planificación y ejecución de la actividad de investigación.	Autores 2 y 3
Recursos	Suministro de materiales de estudio, reactivos, materiales, pacientes, muestras de laboratorio, animales, instrumentación, recursos informáticos u otras herramientas de análisis.	Autores 1, 2 y 3
Software	Programación, desarrollo de software; diseño de programas informáticos; implementación del código informático y de los algoritmos de apoyo; prueba de los componentes de código existentes.	Autores 1, 2 y 3
Supervisión	Responsabilidad de supervisión y liderazgo en la planificación y ejecución de actividades de investigación, incluyendo la tutoría externa al equipo central.	Autores 2 y 3
Validación	Verificación, ya sea como parte de la actividad o por separado, de la replicabilidad/reproducción general de los resultados/experimentos y otros productos de la investigación.	Autores 1, 2 y 3
Visualización	Preparación, creación y/o presentación del trabajo publicado, específicamente la visualización/presentación de datos.	Autores 1, 2 y 3
Redacción / Borrador original	Preparación, creación y/o presentación del trabajo publicado, específicamente la redacción del borrador inicial (incluyendo la traducción sustantiva).	Autora 1
Redacción / Revisión y edición	Preparación, creación y/o presentación del trabajo publicado por los miembros del grupo de investigación original, específicamente revisión crítica, comentario o revisión, incluidas las etapas previas o posteriores a la publicación.	Autores 1, 2 y 3

8. Declaración sobre uso de inteligencia artificial

En este artículo se ha utilizado ChatGPT para una revisión inicial de posibles errores.

9. Referencias bibliográficas

Al-jenaibi, B. y Bulhoon, A. (2024). Investigating adoption determinants, obstacles, and interventions for AI implementation in Emirati media organizations. *South Eastern European Journal of Public Health*, 25(1), 1901-1909. <https://doi.org/10.70135/seejph.vi.2254>

Al-Zoubi, O., Ahmad, N. y Hamid, N. A. (2024). Artificial intelligence in newsrooms: Ethical challenges facing journalists. *Studies in Media and Communication*, 12(1), 401. <https://doi.org/10.11114/smc.v12i1.6587>

Bakke, N. A. y Barland, J. (2022). Disruptive innovations and paradigm shifts in journalism as a business: From advertisers first to readers first and traditional operational models to the AI factory. *SAGE Open*, 12(2), 1-13. <https://doi.org/10.1177/21582440221094819>

Barnett, J. (2023). The ethical implications of generative audio models: A systematic literature review. En *Proceedings of the 2023 AAAI/ACM Conference*

- rence on AI, Ethics, and Society (pp. 146-161). ACM. <https://doi.org/10.1145/3600211.3604686>
- Barrington, S. y Farid, H. (2024). People are poorly equipped to detect AI-powered voice clones. *arXiv*. <https://doi.org/10.48550/arxiv.2410.03791>
- Bayer, J. (2024). Legal implications of using generative AI in the media. *Information & Communications Technology Law*, 33(3), 310-329. <https://doi.org/10.1080/13600834.2024.2352694>
- Beckett, C., Sanguinetti, P. y Palomo, B. (2023). New frontiers of the intelligent journalism. En *Blurring Boundaries of Journalism in Digital Media: New Actors, Models and Practices* (pp. 275-288). Springer International Publishing.
- Benjamins, R. (2020). Towards organizational guidelines for the responsible use of AI. *Frontiers in Artificial Intelligence and Applications*, 325. <https://doi.org/10.3233/FAIA200433>
- Bleyen, V. A., Lindmark, S., Ranaivoson, H. y Ballon, P. (2014). A Typology of Media Innovations: Insights from an Exploratory Study. *The Journal of Media Innovations*, 1(1), 28-51. <https://doi.org/10.5617/jmi.v1i1.800>
- Boczkowski, P. (2004). *Digitizing the news*. MIT Press.
- Brennen, J. S., Howard, P. N. y Nielsen, R. K. (2018). *An industry-led debate: how UK media cover artificial intelligence*. Reuters Institute for the Study of Journalism. <https://www.oxfordmartin.ox.ac.uk/publications/an-industry-led-debate-how-uk-media-cover-artificial-intelligence>
- Broussard, M., Diakopoulos, N., Guzman, A., Abebe, R., Dupagne, M. y Chuan, C. H. (2019). Artificial intelligence and journalism. *Journalism and mass communication quarterly*, 96(3), 673-695. <http://doi.org/10.1177/1077699019859901>
- Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society a Mathematical Physical and Engineering Sciences*, 376(2133), 20180080. <https://doi.org/10.1098/rsta.2018.0080>
- Chan-Olmsted, S. M. (2019). A review of artificial intelligence adoptions in the media industry. *International Journal on Media Management*, 21(3-4), 193-215. <https://doi.org/10.1080/14241277.2019.1695619>
- Danzon-Chamblaud, S. (2021). A systematic review of automated journalism scholarship: guidelines and suggestions for future research. *Open research Europe*, 1(4). <https://doi.org/10.12688/openreseurope.13096.1>
- de-Lima-Santos, M. F. y Ceron, W. (2021). Artificial Intelligence in News Media: Current Perceptions and Future Outlook. *Journalism and Media*, 3(1), 13-26. <https://doi.org/10.3390/journalmedia3010002>
- de-Lima-Santos, M. F., Yeung, W. N. y Dodds, T. (2024). Guiding the way: A comprehensive examination of AI guidelines in global media. *arXiv*. <https://doi.org/10.48550/arxiv.2405.04706>
- Epstein, Z. y Hertzmann, A. (2023). Art and the science of generative AI. *Science*, 380(6650), 1110-1111. <https://doi.org/10.1126/science.adh4451>
- Feiras-Ceide, C., Vaz-Álvarez, M. y Tüñez-López, M. (2022). Artificial intelligence strategies in European public broadcasters: Uses, forecasts and future challenges. *Profesional de la información*, 31(5). <https://doi.org/10.3145/epi.2022.sep.18>
- Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *methadods. Revista de ciencias sociales*, 11(2), m231102a10. <https://doi.org/10.17502/mrcs.v11i2.710>
- García, V., Hernáez, I. y Navas, E. (2022). Evaluation of Tacotron based synthesizers for Spanish and Basque. *Applied Sciences*, 12(3), 1686. <https://doi.org/10.3390/app12031686>
- García-Orosa, B., Canavilhas, J. y Vázquez-Herrero, J. (2023). Algorithms and communication: A systematized literature review. *Comunicar*, 74, 9-21. <https://doi.org/10.3916/C74-2023-01>
- Gondwe, G. (2023). Exploring the Multifaceted Nature of Generative AI in Journalism Studies: A Typology of Scholarly Definitions. *Social Science Research Network*. <https://doi.org/10.2139/ssrn.4410826>
- Gutiérrez-Caneda, B., Lindén, C. y Vázquez-Herrero, J. (2024). Ethics and journalistic challenges in the age of artificial intelligence: Talking with professionals and experts. *Frontiers in Communication*, 9. <https://doi.org/10.3389/fcomm.2024.1465178>
- Gutiérrez-Caneda, B., Vázquez-Herrero, J. y López-García, X. (2023). AI application in journalism: ChatGPT and the uses and risks of an emergent technology. *Profesional de la información*, 32(4). <https://doi.org/10.3145/epi.2023.jul.08>
- Heiselberg, L., Blom, J. N. y van Dalen, A. (2022). Automated news reading in the neural age: Audience reception and perceived credibility. *Journalism Studies*, 23(8), 896-914. <https://doi.org/10.1080/1461670X.2022.2052346>
- Ioscote, F., Gonçalves, A. y Quadros, C. (2024). Artificial Intelligence in Journalism: A Ten-Year Retrospective of Scientific Articles (2014-2023). *Journalism and Media*, 5(3), 873-891. <https://doi.org/10.3390/journalmedia5030056>
- Kim, D. y Kim, S. (2017). Newspaper companies' determinants in adopting robot journalism. *Technological Forecasting and Social Change*, 117, 184-195. <https://doi.org/10.1016/j.techfore.2016.12.002>
- Kumar, V., Kapoor, A., Chaudhary, R. R., Gupta, L. y Khokhar, D. (2024). Preserving integrity: A binary classification approach to unmasking artificially generated voices in the age of deepfakes. En *11th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1449-1454). IEEE. <https://doi.org/10.23919/INDIACom61295.2024.10499177>
- Lewis, D. y Moorkens, J. (2020). A rights-based approach to trustworthy AI in social media. *Social Media + Society*, 6(3). <https://doi.org/10.1177/2056305120954672>
- Liu, Z. (2024). Analysis of the impact of artificial intelligence on the media and film industries. *Lecture Notes in Education Psychology and Public Media*, 35, 219-223.
- Lopezosa, C., Codina, L., Pont-Sorribes, C. y Váñez, M. (2023). Use of generative artificial intelligence in the training of journalists: Challenges, uses and training proposal. *Profesional de la infor-*

- mación, 32(4). <https://doi.org/10.3145/epi.2023.jul.08>
- Naik, V., Mendes, A., Kulkarni, S., Naik, S. y Verlekar, S. P. (2022). Voice cloning in real time. *International Journal for Research in Applied Science and Engineering Technology*, 10(8), 1443-1446.
- Nielsen, R. K. y Ganter, S. A. (2022). The power of platforms shaping media and society. *Oxford Studies in Digital Politics*. <https://reutersinstitute.politics.ox.ac.uk/news/power-platforms>
- Noain-Sánchez, A. (2022). Addressing the impact of artificial intelligence on journalism: The perception of experts, journalists and academics. *Communication & Society*, 35(3), 105-121. <https://doi.org/10.15581/003.35.3.105-121>
- Olsen, R. y Hess, K. (2024). "It's New to Us": Exploring Authentic Innovation in Local News Settings. *Media and Communication*, 12, 7444. <https://doi.org/10.17645/mac.7444>
- Painter, C. (2022). Enhanced Realism or A.I.-Generated illusion? Synthetic voice in the documentary film *Roadrunner*. *Journal of Media Ethics*, 37(4), 296-297. <https://doi.org/10.1080/23736992.2022.2113885>
- Pavlik, J. V. (2013). Innovation and the future of journalism. *Digital Journalism*, 1(2), 181-193. <https://doi-org.ehu.idm.oclc.org/10.1080/21670811.2012.756666>
- Peña-Fernández S., Peña-Alonso U. y Eizmendi-Iraola M. (2023). El discurso de los periodistas sobre el impacto de la inteligencia artificial generativa en la desinformación. *Estudios sobre el mensaje periodístico*, 29(4), 833-841. <https://doi.org/10.5209/esmp.88673>
- Peña-Fernández, S., Meso-Ayerdi, K., Larrondo Ureta, A. y Díaz-Noci, J. (2023). Without journalists, there is no journalism: the social dimension of generative artificial intelligence in the media. *Profesional de la Información*, 32(6). <https://doi.org/10.3145/epi.2023.nov.05>
- Sánchez-García, P., Merayo-Álvarez, N., Calvo-Barbero, C. y Díez-Gracia, A. (2023). Spanish technological development of artificial intelligence applied to journalism: companies and tools for documentation, production and distribution of information. *Profesional de la Información*, 32(3). <https://doi.org/10.3145/epi.2023.may.10>
- Sarrionandia, B., Peña-Fernández, S., Pérez-Dasilva, J. y Larrondo-Ureta, A. (2025). Artificial Intelligence Training in Media: Addressing Technical and Ethical Challenges for Journalists and Media Professionals. *Frontiers in Communication*, 10. <https://doi.org/10.3389/fcomm.2025.1537918>
- Schiff, D. S., Kelley, S. y Camacho Ibáñez, J. (2024). The emergence of artificial intelligence ethics auditing. *Big Data & Society*, 11(4). <https://doi.org/10.1177/20539517241299732>
- Scott, K. M., Ashby, S. y Hanna, J. (2020). "Human, All Too Human": NOAA weather radio and the emotional impact of synthetic voices. En *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-9). ACM. <https://doi.org/10.1145/3313831.3376338>
- Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, 10(10), 1832-1854. <https://doi.org/10.1080/21670811.2022.2063150>
- Sonni, A. F., Hafied, H., Irwanto, I. y Latuheru, R. (2024). Digital Newsroom Transformation: A Systematic Review of the Impact of Artificial Intelligence on Journalistic Practices, News Narratives, and Ethical Challenges. *Journalism and Media*, 5(4), 1554-1570. <https://doi.org/10.3390/journalmedia5040097>
- Terol, T. M. (2023). Innovación mediática: Aplicaciones de la inteligencia artificial en el periodismo en España. *Textual and Visual Media*, 17(1), 41-60.
- Túñez-López, J.-M., Feiras-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>
- Verma, D. (2024). Impact of artificial intelligence on journalism: A comprehensive review of AI in journalism. *Journal of Communication and Management*, 3(02), 150-156. <https://doi.org/10.58966/JCM20243212>
- Wang, S., & Huang, G. (2024). The Impact of Machine Authorship on News Audience Perceptions: A Meta-Analysis of Experimental Studies. *Communication Research*, 51(7), 815-842. <https://doi.org/10.1177/00936502241229794>
- Westlund, O. y Lewis, S. (2014). Agents of Media Innovations: Actors, Actants, and Audiences. *The Journal of Media Innovations*, 1(2), 10-35. <https://doi.org/10.5617/jmi.v1i2.856>
- Wu, S., Tandoc, E. C. y Salmon, C. T. (2019). Journalism reconfigured. *Journalism Studies*, 20(10), 1440-1457. <https://doi.org/10.1080/1461670X.2018.1521299>

Barbara Sarrionandia. Barbara Sarrionandia es licenciada en Periodismo por la Universidad de Navarra y obtuvo el Máster Universitario en Formación del Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanzas de Idiomas, con especialidad en Geografía e Historia, por la Universidad Internacional de La Rioja (UNIR). Ha realizado el curso ejecutivo European Political Economy and Finance impartido por la London School of Economics and Political Science (LSE) y actualmente cursa el programa de Doctorado en Comunicación Social en la UPV/EHU, donde centra su investigación en inteligencia artificial, desinformación y alfabetización mediática. A lo largo de dos décadas, ha desarrollado su trayectoria profesional en medios de comunicación e instituciones tanto a nivel estatal como internacional. ORCID: <https://orcid.org/0009-0006-0053-3730>

Simón Peña-Fernández. Profesor Pleno en el Departamento de Periodismo de la Universidad del País Vasco (UPV/EHU). Sus principales líneas de investigación son el ciberperiodismo, la comunicación digital y la inteligencia artificial. Es el Investigador Principal, junto con Koldobika Meso, del proyecto de investigación «Impacto de la inteligencia artificial y los algoritmos en los cibermedios, los profesionales y las audiencias» (PID2022-138391OB-I00), financiado por el Ministerio de Ciencia e Innovación. ORCID: <https://orcid.org/0000-0003-2080-3241>

Jesús Ángel Pérez-Dasilva. Profesor Pleno en el Departamento de Periodismo de la Universidad del País Vasco (UPV/EHU). Sus principales líneas de investigación son el ciberperiodismo, la comunicación social, las redes Sociales y la innovación social. Es el Investigador Principal del Proyecto «Impacto de la inteligencia artificial en los medios de comunicación vascos y sus profesionales» (US 23/10), financiado por la UPV/EHU. ORCID: <https://orcid.org/0000-0002-3383-4859>

Kazetarienezako tresnak adimen artifizialak indartutako desinformazioaren aurka

(Tools for journalists against AI-driven disinformation)

Sarrionandia, Barbara

Euskal Herriko Unibertsitatea (UPV/EHU), Gizarte eta Komunikazio
Zientzien Fakultatea, Kazetaritza Saila. Sarriena z/g. 48940 Leioa.
bsarrionandia001@ikasle.ehu.eus

Recep.: 18.09.2024

BIBLID [eISSN 1988-3935 (2024), 20; 111-138]

Acep.: 16.11.2024

Kazetaritzan adimen artifiziala (AA) berraztertzeak guztiz eraldatzen du informazioa sortzeko, banatzeko eta kontsumitzeko era, baina desinformazioarekin eta etika profesionalarekin lotutako arrisku esanguratsuak ere badakartza. Artikulu honek aztertzen du nola tresna aurreratuek, besteak beste eredu sortzaileek eta deepfake-ek, aldatzen dituzten ekosistema informatiboak eta handitzen duten eduki faltsuen hedapena. Literatura zientifikoaren azterketan eta AAren eta desinformazioaren arteko harremanari buruzko ikerketa berrietan oinarrituta, kazetaritzak aurrez aurre dituen erroka etiko eta operatibo nagusiak identifikatzen dira. Gainera, hainbat irtenbide aztertzen dira, hala nola alfabetatze mediatikoa, fact-checking-a eta teknologia horien erabilera arduratsua bultzatzen duten erregulazio digitaleko politiken diseinua. Lan honek berrikuntza teknologikoa, prestakuntza etikoa eta nazioarteko lankidetzaren uztartuko dituen diziplina anitzeko ikuspegiaren garrantzia azpimarratzen du, kazetaritzan konfiantza indartzeko eta gero eta konplexuagoa den ingurune digital batean informaziozko eskubidea babesteko.

Gako hitzak: Kazetaritza. Adimen artifiziala (AA). Desinformazioa. Alfabetatze mediatikoa. Fact-checking-a.

La integración de la inteligencia artificial (IA) en el periodismo está transformando radicalmente la forma en que se produce, distribuye y consume la información, pero también plantea riesgos significativos relacionados con la desinformación y la ética profesional. Este artículo analiza cómo herramientas avanzadas, como los modelos generativos y los deepfakes, están alterando los ecosistemas informativos y amplificando la propagación de contenido falso. Basado en un análisis de literatura científica y en estudios recientes sobre la relación entre IA y desinformación, se identifican los principales retos éticos y operativos que enfrenta el sector. Además, se exploran soluciones como la alfabetización mediática, el fact-checking y el diseño de políticas de regulación digital que promuevan un uso responsable de estas tecnologías.

Sarrionandia, B: Kazetarientzako tresnak adimen artifizialak indartutako desinformazioaren...

Este trabajo enfatiza la importancia de un enfoque multidisciplinario que combine innovación tecnológica, formación ética y colaboración internacional para fortalecer la confianza en el periodismo y proteger el derecho a la información en un entorno digital cada vez más complejo

Palabras clave: Periodismo. Inteligencia artificial (IA). Desinformación. Alfabetización mediática. Fact-checking.

The integration of artificial intelligence (AI) into journalism is radically transforming the ways in which information is produced, distributed, and consumed. However, it also presents significant risks concerning disinformation and professional ethics. This article examines how advanced tools—such as generative models and deepfakes—are reshaping the information ecosystem and amplifying the spread of false content. Drawing on a review of academic literature and recent studies on the relationship between AI and disinformation, the article identifies the main ethical and operational challenges facing the sector. It also explores potential solutions, including media literacy, fact-checking, and the development of digital regulation policies aimed at fostering the responsible use of these technologies. The study highlights the importance of a multidisciplinary approach that combines technological innovation, ethical training, and international cooperation to reinforce trust in journalism and safeguard the right to information in an increasingly complex digital environment.

Keywords: Journalism. Artificial intelligence (AI). Disinformation. Media literacy. Fact-checking.

L'intégration de l'intelligence artificielle (IA) dans le journalisme transforme radicalement les modes de production, de diffusion et de consommation de l'information. Elle soulève néanmoins d'importants risques liés à la désinformation et à l'éthique professionnelle. Cet article analyse comment des outils avancés – tels que les modèles génératifs et les deepfakes – bouleversent les écosystèmes informationnels et accentuent la propagation de contenus faux. En s'appuyant sur une revue de la littérature scientifique et sur des études récentes concernant le lien entre IA et désinformation, il identifie les principaux défis éthiques et opérationnels auxquels le secteur est confronté. Sont également explorées des pistes de solution telles que l'éducation aux médias, le fact-checking, ainsi que l'élaboration de politiques de régulation numérique favorisant un usage responsable de ces technologies. Ce travail souligne l'importance d'une approche multidisciplinaire, alliant innovation technologique, formation éthique et coopération internationale, afin de renforcer la confiance dans le journalisme et de préserver le droit à l'information dans un environnement numérique de plus en plus complexe.

Mots clés: Journalisme. Intelligence artificielle (IA). Désinformation. Éducation aux médias. Vérification des faits.

1. SARRERA

Kazetaritzaren egungo egoeran, AAre eta desinformazioaren arteko konbergentziak erronka-sare konplexua sortu du komunikazioko profesionalentzat. Fenomeno horrek informazioa ekoizteko eta zabaltzeko dinamikak birdefinitzeaz gain, kontu etiko eta sozialak ere planteatzen ditu hedabideekiko konfiantzari, informazioaren egiazkotasunari eta audientzien manipulazioari buruzkoak. Artikulu honetan, aztertzen da tresna teknologikoak implementatzeak eta desinformazio-estrategiak ugaritzeak nola eraldatzen dituzten kazetaritza-praktikak eta, aldi berean, herritarrek informazio fidagarria eskuratzeko dituzten erak (Jankowski, 2018; Kapantai et al., 2021; García-Marín & Salvat-Martinrey, 2022; Pérez-Escobar, Lilleker & Tapia-Frade, 2023).

Lehenik eta behin, egia-ostearen aroan bizi gara, non iritzi publikoak eratzean emozioak gertakariei gailentzen zaizkien. Testuinguru horrek, gainera, albiste faltsuak, konbentzitzeko edo manipulatzeko diseinatuak, azkarrago hedatzea erraztu du, baita sare sozialetan albiste egiaztatuek baino audientzia handiagoa lortzea ere (Vosoughi, Roy & Aral, 2018; López-Pan & Rodríguez-Rodríguez, 2020). Hori dela eta, Osasunaren Mundu Erakundeak, arazoaren larritasunaz jabetuta, «infodemia» terminoa sortu du informazio faltsuaren ugaritze masiboa deskribatzeko, bereziki mundu-mailako krisi egoeretan, hala nola COVID-19aren pandemian (OME, 2020).

Hala eta guztiz ere, desinformazioak sustrai historikoak dituen arren —Sobietar Batasunaren propaganda-estrategietatik hasi eta kanpaina politiko garaikideetaraino—, argi dago teknologia aurreratuak erabiltzeak —*deepfake*-ak eta hizkuntzaren eredu sortzaileak kasu— hedadura eta eraginkortasuna handitu dituela. Pascual-Presaren (et al. 2024) ustez, tresna horiek eduki engainagarria masiboki sortzea sustatzeaz gain, errealitatearen eta faltsutasunaren arteko bereizketa ere zailtzen dute, eta horrek zuzenean eragiten du publikoak komunikabide tradizionaletan duen konfiantzan. Alde horretatik, Europako Batzordeak (2018) honela definitzen du desinformazioa: «informazio egiaztagarriki faltsua edo engainagarria, irabazi ekonomikoak lortzeko edo asmo txarrez manipulatzeko sortua, aurkeztua eta zabaldua». Definizio horrek, gainera, borrokarako politiken eta estrategien diseinua oinarritu du, hala nola Desinformazioaren aurkako Ekintza Plana eta SOMA, PROVENANCE eta EUNOMIA bezalako lankidetzaproiektuak.

Bestalde, zenbait ikerketek desinformazioa hiru mota nagusitan sailkatu dute: *misinformation*, *disinformation* eta *malinformation* (Wardle & Derakhshan, 2017). Kategoria horiek barne hartzen dituzte bai errore zintzoak bai nahita egindako eragin handiko manipulazioak eta fenomenoak eremu digitalean duen aniztasuna eta konplexutasuna islatzen dute. Ondorioz, sare sozialetako informazio faltsuaren biralizazioa —plataforma horien algoritmiko-egiturak elikatua—, manipulazio masiborako tresna ahaltsua da eta iturri fidagarrien identifikazioa zailtzen du eta polarizazio soziala sustatu (Costa-Sánchez & López-García, 2020; Montemayor-Rodríguez & García-Jiménez, 2021).

Era berean, AAn egindako aurrerapenak arazoa areagotu dute. ChatGPT gisako teknologiek eta sare neuronaletan oinarritutako beste tresna batzuek erakutsi dute mehatxua izan daitezkeela, baina aukera ere bai. Alde batetik, eduki koherentea eta konbentzigarria sortzeko duten gaitasunak ezin hobeak egiten ditu informazio faltsua hedatzeko; bestetik, desinformazio-ereduak identifikatzeko eta egiaztapen-lana indartzeko ere erabil daitezke (Baloch et al., 2024; Pascual-Presa et al., 2024). Hala ere, Taplinek (2023) adierazten duen moduan, teknologia horiekiko mendekotasun gero eta handiagoak arrisku etikoak sortzen ditu horien erabileraren eta erregulazioaren inguruan.

Gainera, AA erredakzioetan txertatzeak aukera berriak ireki ditu, baina alfabetatze mediatikoa eta trebakuntza profesionala indartuko dituzten politiken beharra ere areagotu du. García-Galera-ren, Del-Hoyo-Hurtado-ren eta Blanco-Alfonso-ren arabera (2020), kazetariak erronka bati egin behar diote aurre: asmatutako albisteak, zehaztasunik gabekoak edo okerrak bereizteari, informazioaren ugaritasuna eta sare sozialek lehen albiste-iturri gisa duten protagonismoa ezaugarri dituen ingurune digital batean lan egiten duten bitartean (Newman, 2022).

Artikulu honen asmoa AAK eta desinformazioak kazetaritzari nola eragiten dioten aztertzea da, etika profesionalaren, teknologiaren eta desinformazioaren aurka borrokatzeko estrategien arteko elkarguneak aztertuko dituen ikuspegia proposatuz. Horrela, esparru multidimentsional bat eskaini nahi da komunikazioko profesionalei aukera emango diena informazio-errealitate berri horrek planteatzen dituen erronkei eta aukerei heltzeko.

1.METODOLOGIA

Ikerketa honek AAK kazetaritzan duen eragina eta desinformazioarekin duen harremana aztertzen ditu, alderdi etikoak, teknologikoak eta sozialak uztartzen dituen diziplinarteko ikuspegi batetik. Era honetan, alfabetatze mediatikoarekin, *fakenews*-ekin eta desinformazioaren aurka borrokatzeko estrategiekin lotutako kontzeptu nagusiak aztertzen dira, arreta jarritz problematika horiek kazetaritzan duten ondorioetan. Horretarako, hainbat datu-base akademiko eta iturri garrantzitsu erabili dira, besteak beste, ikerketa berriak eta gaiari buruzko aipamen giltzarriak.

Testuinguru horretan, ikuspegi metodologikoa elkarri lotutako hiru ikerketa-galderaren inguruan egituratu da:

1. Nola ari da AA kazetaritzaren dinamikak aldatzen?
2. Zein dira AAeko teknologiak erabiltzearekin lotutako arrisku nagusiak albisteak sortu eta zabaltzean?

3. Zer tresna eta estrategia daude teknologia horiek hedatutako desinformazioari aurre egiteko?

Analisiaren zorroztasuna bermatzeko, ikuspegi kualitatibo bat erabili da eta horri esker identifikatu ahal izan dira AAre erabileraren ereduak eta joerak kazetaritzaren esparruan. Horrek barne hartu du desinformazioari lotutako arriskuen kategorizazioa eta gobernuek, nazioarteko erakundeek eta komunikabideek ezarritako estrategien ebaluazioa.

Azkenik, aztertutako ikuspegiak artikulatu eta esparru ulerkor bat eskaini nahi izan da, kazetariei, akademikoei eta legegileei aukera emango diena AAK desinformazioaren aroan planteatzen dituen erronkak ulertzeko eta horiei heltzeko. Hala, ikerketak oinarria eskaintzen du berrikuntza teknologikoa, alfabetatze mediatikoa eta etika profesionala uztartzen dituzten konponbide integralak aztertzeko.

3. AZTERKETA

3.1 Kazetaritzaren bilakaera: aurrerapen teknologikoetara eta erronka berrietara egokitzea

Kazetaritzaren esparrua etengabe eraldatu da historikoki teknologia emergenteak txertatu ahala. Azken hamarkadetan ordea, aldaketa hori disruptiboa izan da, batez ere AA erredakzioetan inplementatu denean. Fenomeno horrek "exokazetaritza" (Tejedor & Vila, 2021), "robot kazetaritza", "kazetaritza konputazionala", "kazetaritza algoritmikoa" eta "kazetaritza automatizaturik" bezalako terminoak sortu ditu, bilakaeran dagoen eta oraindik definizio argi eta zehatzik ez duen eztabaida islatuz (Mooshammer, 2022).

Kazetaritza-lanen automatizazio gero eta handiagoak prozesuak optimizatzeaz gain, kezka etikoak eta sozialak ere planteatu ditu. Zorroztasunaren galera eta informazioa egiaztatzearen ahultzea arazo nagusi bihurtu dira egungo demokrazientzat (López-Martín, Gómez-Calderón & Córdoba-Cabús, 2023). Fenomeno hori, gainera, larriagotu egin da ohiko hedabideek informazioa egiaztatu beharrean eduki desinformatiboak erreproduzitzeko joera areagotu dutelako. Horren ondorioz, infoxikazio-testuingurua sortzen da eta informazio-erakundeekiko konfiantza galtzen da (López-Martín et al., 2023; Salaverria et al., 2020).

Testuinguru horretan, kazetariaren rola «gatekeeper» izatetik «gatewatcher» izatera igaro da (Bruns, 2003) eta funtzio kritikoa hartu du edukien zaintzan. Trantsizio hori bereziki garrantzitsua da gerra-gatazken testuinguruetan adibidez; izan ere, kasu horietan, AAK sortutako desinformazioa kontrolik gabe zabaltzen da eta komunikabideen sinesgarritasunari eragiten dio. Duela gutxiko adibideek, hala nola Ukrainako gerrak edo Gazako krisiak, frogatu dute

kazetariak informazioa hautatu, egiaztatu eta testuinguruan kokatu behar dutela narratiba faltsuak indargabetzeko, «eduki-zaintzaile» gisa jardunez (Loo-Vázquez et al., 2016).

Alde horretatik, teknologia funtsezko osagaia bihurtu da komunikabideen estrategian eta bilakaeran (Vállez & Codina, 2018). Hala ere, erredakzioetan txertatzearekin batera ez dator beti komunikazioko profesionalen prestakuntza akademiko egokiarekin batera eta horrek mugatu egiten du egungo erronketara egokitze gaitasuna (Ufarte-Ruiz, Fieiras-Ceide & Tüñez-López, 2020). Hala eta guztiz ere, kazetaritza automatizatua nagusitu egin da bai praktika profesionalean bai literatura akademikoan, fenomenoaren ñabardurak islatzen dituzten izen ugariekin (Cohen, Hamilton & Turner, 2011; Karlsen & Stavelin, 2013).

Erredakzioetan AA sortzailearen tresnak egoteak erronkak planteatzen ditu, esaterako kazetaritzaren rol sozialaren balizko gainbehera eta informazioaren egiazkotasunarekiko gero eta kezka handiagoa. Kim et al. (2021) eta López-Jiménez & Ouariachi (2020) autoreek azpimarratzen dute teknologia honek lanbidearen oinarriak birpentsatzera behartzen duela, bereziki makinek oso-osorik sortutako edukien hazkundearen aurrean. Kasu batzuetan, Europako eta Japoniako hedabideek albiste automatizatuak argitaratzen dituzten «hedabide sintetikoen» ereduak egin dituzte eta joera horrek, eraginkorra izan arren, galdera larriak planteatzen ditu informazioaren egiazkotasunari eta fidagarritasunari buruz (Ufarte-Ruiz, Murtzia-Verdú & Tüñez-López, 2023).

AArekiko gero eta mendekotasun handiagoak erregulazio eta kontrol profesional handiagoaren beharra dakar. Smuharen arabera (2021), printzipio editorialak egokitu egin behar dira teknologia horien erabileran gardentasuna eta etika bermatzeko. Smuharen arabera (2021), printzipio editorialak egokitu egin behar dira teknologia horien erabileran gardentasuna eta etika bermatzeko. Wu-k, Tandoc-ek eta Salmon-ek (2019) ideia hori indartzen dute, argudiatzen baitute kazetaritzak egiazkotasun- eta erantzukizun-balioak lehenesten jarraitu behar duela, baita tresna automatizatuak erabiltzen direnean ere. Gainera, hainbat ekimen proposatu dira, hala nola testuen egiletza nahitaez adieraztea eta hala, irakurleei aukera ematen zaie eduki hori gizakiek edo makinek sortu duten identifikatzeko, irakurketa informatuagoa eta kritikoagoa izan dadin.

Bestalde, gaur egungo ekosistema mediatikoak logika klasikoak eta digitalak bateratzen ditu eta fenomeno horrek kontsumo eta ekoizpen informatiboko dinamikak eraldatzen ditu. Kazetaritzaren muga tradizionalen desitxuratze horrek sekulako lehia sortu du hainbat eragileren artean publikoaren arreta erakartzeko eta erronka hori Carlson-ek eta Lewis-ek (2015), Chadwick-ek (2013) eta Van-Aelst-ek et al. (2017) deskribatzen dute audientzien leialtasun zatikatuaren borroka gisa.

Agertoki berri horretara egokitzeko, aurrerapen teknologikoak ez ezik, gogoeta sakona ere egin beharko da gizarte garaikidean kazetaritzak duen eginkizunari buruz. AA eta kazetaritza bateratzeak erronka konplexuak planteatzen dituen arren, lanbidearen etorkizuna birpentsatzeko aukerak ere zabaltzen ditu, betiere lan hori historikoki definitu duten printzipio etikoei eusten bazaie.

3.1.1. AAre bilakaera kazetaritzan: hastapenetatik gaur egungo erabileretara

Kazetaritzaren esparruan AAre historia duela zenbait hamarkadako da, hainbat garaitako lorpen tekniko eta kontzeptualak bateratzen dituzten sustraiekin. «Adimen artifiziala» terminoa 1956an erabili zuen lehen aldiz publikoki John McCarthyk Dartmoutheko Biltzarrean, bere ikerketaren hasiera formala diziplina gisa markatuz (Russell & Norvig, 2022). Hala ere, haren oinarriak Alan Turing-ek Bigarren Mundu Gerran egindako aurrerapenekin lotzen dira, bereziki mezuen deskodetzeari dagokionez, baita 1840ko hamarkadan Lady Ada Lovelace-k zituen etorkizun-seneko ideiekin ere, zeinak makina programagarrien etorkizuneko inplikazioak aurreratu zituen (Boden, 2022).

AAre aurrerapenak bulkada nabarmena izan zuen laurogeiko hamarkadan, batik bat problema matematikoak ebazteko aplikazioan eta testu eleaniztunen azterketan, eta horrek etorkizuneko ezarpenetarako oinarriak ezarri zituen (Ufarte-Ruiz & Manfredi-Sánchez, 2019). Hasierako aurrerapauso horiek gorabehera, 2000ko hamarkadaren erdialdera arte ez zen nabarmen zabaldu esparru akademiko eta industrialetan eta, horri esker, AAKo teknologiek kazetaritzan ezartzeko moduak aldatu zituzten sofistikazio-mailak lortu zituzten (Crawford, 2021).

Ezarpen praktikoei dagokienez, albisteen ekoizpen automatizatuaren lehen esperimentuak 2010ekoak dira, The Big Ten Network-en proiektukoak, zeinak kirol-erreportajeak sortzeko automatizazio-sistema bat txertatu zuen (Canavilhas, 2022). Nahiz eta ez zen izan testu-edukia sortzeko AA erabiltzen zen lehen aldia, proiektu horrek inflexio-puntu bat markatu zuen industria mediatikoan AA txertatzeko orduan. Geroago, 2014an, Los Angeles Times-en Quakebot softwarearen garapenak lurrikaren berri ematea automatizatu zuen eta, horrela, AAre erabilera finkatu zen albiste espezializatuen ekoizpenean (Sánchez-García et al., 2023).

Teknologiak aurrera egin ahala, Associated Press eta Le Monde bezalako erakundeek AA txertatu zuten haien irismena handitzeko eta prozesuak optimizatzeko. Associated Press albisteak sortzeko sistema automatizatuak erabiltzen hasi zen eta Le Monde, berriz, Data2Content teknologia erabili zuen hauteskundeetako emaitzei buruzko notizia laburrak ekoizteko, tresna horiek hainbat informazio-testuingurutara nola egokitu daitezkeen erakutsiz (Lichterman, 2017; Sánchez-González & Sánchez-González, 2017). Aldi berean, Narrative Science eta Automated Insights enpresek berebiziko

garrantzia izan zuten hizkuntza naturala prozesatzeko eta sortzeko teknologien garapenean eta kazetaritza-eduki automatizatua sortzea ahalbidetu zuten (Sánchez-García et al., 2023). Espainian, Vocento MediaLab-en Medusa gisako proiektuek AAren aplikazioa aztertu zuten eremu espezifikoetan, hala nola hondartzen eta eski-pisten egoeraren monitorizazioan (Ufarte-Ruiz & Manfredi-Sánchez, 2019).

Duela gutxi, AppliedXL eta The Newsroom bezalako ekimenek automatizazioa hurrengo mailara eramanez, datuak kazetaritza-informazio bihurtuz denbora errealean (Adami, 2023). Adibide berritzaileak dira El Confidencial-en AnaFut bot-a, zeinak kirol-kronika automatizatuak sortzen dituen (Rojas-Torrijos & Toural-Bran, 2019), eta Narrativa enpresa berriaren Gabriele softwarea, Espainiako hainbat komunikabiderekin elkarlanean kazetaritza-testuak sortzeko diseinatuak (Ufarte-Ruiz & Manfredi-Sánchez, 2019; Sánchez-García et al., 2023). Tresna horiek AAk albisteen ekoizpena optimizatzeko duen ahalmena frogatzeaz gain, idazketa modernoetan teknologia automatizatuekiko gero eta mendekotasun handiagoa ere nabarmentzen dute.

Etengabeko eboluzioaren panorama honetan, 2022ko azaroan ChatGPT merkaturatzea mugari garrantzitsua izan zen erabilgarri jarri zen lehen astean milioi bat erabiltzaile baino gehiago atzitu zituelako (Altman, 2022; Roose, 2022; Lock, 2022). Datu kopuru handietan oinarritutako eredu honek giza interakzioak simulatzen ditu eta hainbat arlotan da erabilgarria, arlo akademikoan laguntzen, aipuak sortzen eta idazteko trebetasunaren atzerarikaduran besteak beste (Aljanabi et al., 2023; Aydere & Karaarslan, 2022; Gilat & Cole, 2023; Golan et al., 2023; Huang & Tan, 2023). Zeregin konplexuak sinplifikatzeko eta informaziorako sarbidea errazteko duen gaitasunak kazetaritza garaikidean AA erabiltzeari dagozkion aukerak zein erronkak islatzen ditu.

Beraz, AAk kazetaritzan izan duen bilakaerak informazioa nola sortzen eta kontsumitzen den birdefinitzeaz gain, kazetariak ekosistema mediatiko hibrido batean duten zeregina birplanteatzen du. Aurrerapauso horiek aukera esanguratsuak eskaintzen badituzte ere, erronka etiko eta profesionalak ere badituzte eta haiei heldu behar zaie kazetaritzaren egiazkotasuna, fidagarritasuna eta erantzukizuna bermatzeko.

3.2. Erronka etikoak eta joerak kazetaritzan AAren aurrean

AA kazetaritzan sartzeak iraultza sakona ekarri du eta aurrekaririk gabeko aukerak sortu ditu, entzungor egin ezin zaien erronka etiko eta profesionalak planteatzen dituen bitartean. Kazetaritza-funtzio oinarritzkoetan giza lanaren ordezkariak erabiltzeak eztabaidak piztu ditu teknologia horiek gizartearen eraginaren inguruan. Dilema horrek profil profesionalak doitzeko ez ezik, kazetaritza-graduetako programa akademikoak birformulatzea ere eskatzen du. Beraz, prestakuntza etiko-deontologikoa funtsezkoa bihurtzen

da kazetari berriak Aaren erabileran trebatzeko, pentsamendu kritiko eta independenterako gaitasuna galdu gabe (Peña-Fernández, Meso-Ayerdi & Larrondo-Ureta, 2023).

Ekosistema digital konplexuan, informazio-bolumenaren hazkundeak informazio-desordenen berezko ahultasunak areagotu ditu (Wardle & Derakhshan, 2017). Fenomeno horrek, informazio-aniztasuna murrizteaz gain, polarizazio sozial eta politikoa ere sakonago egin du eta audientziak gero eta eduki faltsu gehiagoren eraginpean jarri ditu (Marwick & Lewis, 2017; Sunstein, 2017). Datuen erabileraren etika, edukien egiazkotasuna eta informazio-prozesuen gardentasuna, ondorioz, oinarritzko kezka bihurtu dira (Hansen et al., 2017; Brundage et al., 2018; Pihlajarinne & Alén-Savikko, 2022).

Egoera horrek garrantzi berezia hartzen du komunikazio politikoaren eremuan, sare sozialek informazio-fluxua eten baitute, konbentzitzeko tresna bihurtu baitira eta, aldi berean, manipulaziorako eremu egokia ere bai. Horrek iritzi publikoa kutsatu du igorleen borondatea islatzen duten narratiben bidez, botoaren polarizazioan eta osasun demokratikoan eragin zuzena izanik (Enguix, 2020; Casetxe-Ripollés, Doménech-Fabregat & Alonso-Muñoz, 2023). Sare sozialetako esposizio selektiboak informazio-burbuilak eta oihartzun ganberak sortzen lagundu du, zeinek erabiltzaileei ikuspegi ezberdinetarako sarbidea mugatzen dieten eta ideologia itxiak indartzen dituzten (Pariser, 2011; Cinelli et al., 2021; Masip, Suau & Ruiz-Caballero, 2020). Gainera, desinformazio masiboa zabaltzeko diseinatutako *socialbots* eta *softwareen* erabilerak konplexutasun-geruza bat gehitu dio arazoari, iritzi publikoa moldatzen duten narratiba faltsuak sendotuz (Boshmaf et al., 2011; Rossetti & Zaman, 2023).

Konplexutasun horren adibide bat *microtargeting*-a da, edukia pertsonalizatzeko audientzien ohiturak eta lehentasunak aztertzen dituen AAn oinarritutako estrategia. Teknika horrek komunikazio zuzenagoa ahalbidetzen badu ere, erronka etikoak ere mahaigaineratzen ditu, erabaki algoritmikoak diskurtso publikoaren eraketarekin konbinatzen baititu, kazetaritzaren objektibotasunean eraginez eta albiste faltsuak zabaltzea bultzatuz (Barbu, 2014; Gillespie, 2010; Bradshaw et al., 2020). Egiaztapen- eta gezurtapen-estrategiak, beharrezkoak izan arren, mugatuak izan ohi dira; izan ere, haien ondorioak, nagusiki, jada sinesmen jakin batzuk partekatzen dituzten audientzietara mugatzen dira (Van-der-Meer & Hameleers, 2019).

Eragin gehien duten Aaren tresnen artean, ChatGPT nabarmendu da edukia sortzeko eta giza interakzioak simulatzeko duen gaitasunagatik. 2022ko azaroan kaleratu zenetik alabaina, kritikak jaso ditu informazio okerra zabaltzeagatik. ChatGPT-3.5 bertsioa «desinformazioaren hedatzaile handizat» jo zen (Brewster, Arvanitis & Sadeghi, 2023), eta ChatGPT-4 bertsioak, berriz, hobekuntzak ekarri zituen arren, iturriak aipatzea adibidez, oraindik ere erronka nabarmenak ditu. NewsGuard-en azken ebaluazioek erakutsi dute okerreko informazioaren hedapenak bere horretan jarraitzen duela,

eta, gainera, aurreko bertsioetan baino ohikoagoa eta sinesgarriagoa dela (Brewster et al., 2023).

AA kazetaritzan txertatzeak beste enpresa teknologiko batzuk antzeko tresnak garatzera bultzatu ditu. Googlek Bard aurkeztu zuen, ChatGPTren antzeko elkarriketa-eredua eta Microsoftek, berriz, AA generatiboaren gaitasunak txertatu zituen bere bilatzailean, OpenAI-rekin lankidetzan (Google, 2023; Magallón-Rosa, Fernández-Castrillo & Garriga, 2023). Elon Muskek berak ere AAKo bere enpresa merkaturatu du (X.AI), teknologia horiek ekosistema mediatikoan dituzten aukerak zabalduz (Prego, 2023).

Arrisku etikoak gorabehera, plagioa egiteko aukera eta gardentasun akademikorik eza barne (Alkaissi & McFarlane, 2023; Alshater, 2022), kazetarien eta akademikoen pertzepzio orokorra da AAK gaitasun profesionalak nabarmen hobetu ditzakeela. Tresna horiek ohiko zereginetan denbora aurrezteaz gain, kazetariei aukera ematen diete jarduera kognitiboagoetan kontzentratzeko, esaterako ikerketan, sormenean eta analisi kritikoan (Wu, Tandoc & Salmon, 2019; Guzman & Lewis, 2020; Aljanabi et al., 2023).

Praktikan, AA dagoeneko erabiltzen da datuak aztertzeko eta joerak iragartzeko. Virality Oracle (The Washington Post-ek erabiltzen duena) eta Lynx Insights (Reuters-ena) tresnek, adibidez, erakusten dute nola teknologia horiek optimizatu dezaketen albisteak sortzeko prozesua eta nola hobetu dezaketen haien irismena (García-Avilés, 2019; Agarwal, 2018). Hala ere, hor jarraitzen dute AAK lan-prekarizazioan eta independentzia editorialean duen eraginari buruzko kezkek, funtsezko elementuak baitira kazetaritzan erakunde gisa konfiantza mantentzeko (Manfredi-Sánchez & Ufarte-Ruiz, 2020; Peña-Fernández et al., 2023).

Beraz, kazetaritzan AA txertatzeak erregulazio sendoa eta uztartzen zaizkion arriskuak arinduko dituen ikuspegi etikoa eskatzen du, batez ere desinformazioari dagokionez. AAK sektorea modu positiboan eralda dezakeen arren, eraldaketa hori printzipio editorialekiko konpromisoarekin batera joan behar da, bai eta paradigma berri horren erronkei aurre egiteko kazetarien trebakuntzarekin batera ere.

3.3. Desinformazioak eta AAK bultzatutako informazio- eta konfiantza-krisiak

AAren teknologien gorakadak informazio-panorama eraldatu du, baina errorraka garrantzitsuak planteatuz. Besteak beste, bilaketa-emaitzetan AAK sortutako erantzunak sartzea nabarmentzen da, Googlek 2023ko maiatzean iragarritakoa adibidez (Hazard-Owen, 2023). Horri erantzunez,

OpenAIk ChatGPTren ordainpeko bertsio berria agindu zuen, funtzio hobetuez hornitua, sarerako zuzeneko konexioa barne, zeina lehen kanpoko softwarearen menpe zegoen.

Hala ere, kazetaritzaren testuinguruan, ChatGPTk muga esanguratsuak ditu. Haren ezagutza estatikoak, 2021era arte bildutako datuetan oinarrituak, ematen duen informazioaren gaurkotasunari eta garrantziari buruzko galderak planteatzen ditu, batez ere etengabe eboluzionatzen ari den albisteen ingurunean. Alderdi horrek azpimarratzen du tresna horiek modu kritikoan eta gainbegiratuan erabili behar direla informazio-esparruan.

Aldi berean, aro digitalak desinformazioaren ondorioak areagotu ditu eta fenomeno horrek gizartean eragin sakona du. MITen ikerketen arabera, albiste faltsuek % 70 aukera gehiago dituzte partekatuak izateko eta biralak bihurtzeko egiazko informazioarekin alderatuta (Vosoughi *et al.*, 2018). Portaera hori bultzatu egiten da audientziek «Emirec» (Cloutier, 1973) gisa jarduteko duten gaitasunagatik, igorleen eta hartzaileen rola txandakatuz, eta horrek erraztu egiten du eduki sentsazionalistaren hautaketa eta hedapena, askotan elkarrekintza sozialak maximizatzeko diseinatuak.

Era berean, AAI tresna indartsuak eskaini ditu desinformazioa eskala handian sortu eta zabaltzeko, eta horrek iritzi publikoa manipulatzeko eta gizarte-egiturak ezegonkortzeko mehatxua dakar. Duela gutxiko adibideak dira, besteak beste, COVID-19aren pandemian konspirazio-teoriak ugartzea, klima-alaketari buruzko desinformazioarekin eta teoria terraplanistekin batera (Rodríguez & Giri, 2021; León, López-Goñi & Salaverría, 2022). Fenomeno hau ez da isolatua; izan ere, Cambridge Analytica bezalako kasuek eta Estatu Batuetako hauteskunde presidentzialetako interferentzia errusiarrak agerian utzi dute desinformazio-kanpainen gizarte-konfiantza apaldu eta zatiketarako areagotu ditzaketela (Colley, Granelli & Althuis, 2020; Casero-Ripollés, Doménech-Fabregat & Alonso-Muñoz, 2023).

Audientzien rola funtsezkoa da desinformazioaren biralizazioan. Starbird-ek, Arif-ek eta Wilson-ek (2019) barneratutako «unwilling crowds» kontzeptuak azaltzen du pertsonen nola partekatzen eta moldatzen dituzten beren emozio edo sinesmen sakonetara iristen zaizkien edukiak. Gainera, Chadwick-en, Vaccari-ren eta O'Loughlin-en (2018) ikerketek azpimarratzen dute prentsa sentsazionalistaren irakurleek eduki engainagarria zabaltzeko joera handiagoa dutela, oihartzun ganberen eta informazio-burbuilen ondorioak areagotuz (Pariser, 2011; Cinelli *et al.*, 2021).

Hedabideek desinformazioari aurre egiteko ahaleginak gorabehera, haren eragin anplifikatzaileak agerikoa izaten jarraitzen du. Ikerketek erakutsi dute komunikabideek berrietan ohartarazpenak edo zuzenketak egiten dituztenean ere, mezu zuzentzaile horiek ez direla nahikoa informazio faltsuaren hedapena indargabetzeko (Lewandowsky *et al.*, 2012; Walter *et al.*, 2020; Waisbord, 2018). Donald Trump 2016an boterera iritsi izanak inflexio-puntu bat markatu

zuen fenomeno honetan, albiste faltsuen azterketa areagotuz esfera publikoko elementu nagusi gisa (Orbegozo-Terradillos, Morales-i-Gras & Larrondo-Ureta, 2020; Salaverria et al., 2020), eta errepikatu egin da beste behin 2024an ere, Donald Trumpen garaipenarekin.

Hala eta guztiz, herritarrek oraindik ere uste dute hedabideek eta eragile politikoek ez dutela behar adina egiten arazo horri aurre egiteko. Horrek erakunde askorengan ondorioak dituen konfiantza-krisia eragin du, komunikabideen enpresak, alderdi politikoak eta GKEak barne (Edelman, 2023; Pérez-Latre, 2022). Komunikazio publikoaren esparruan, mesfidantza hori zuzenean islatzen da informazio-kontsumoan; izan ere, konfiantza mediatikoak nabarmen eragiten du ikus-entzuleek zer berri kontsumitu erabakitzeke orduan (Fletcher & Park, 2017; Tsfaty & Ariely, 2014).

Konfiantza mediatikoa, literatura akademikoan asko aztertua (Kohring & Matthes, 2007; Meyer, 1988; Otto & Köhler, 2018; Strömbäck et al., 2020), informazio-produktuaren kalitatearen eta haren zehaztasuna bermatzeko erabilitako prozeduren arabera da. Arestiko ikerketek nabarmendu dute ingurune mediatikoko egiturazko tentsioek erakunde demokratikoak ahultzen laguntzen dutela, desinformazioa sakabanatzea erraztuz eta hedabideekiko konfiantza are gehiago higituz (Bennet & Livingstone, 2018; Fawzi et al., 2021).

Azken finean, desinformazioak arrakala ideologikoak handiagotzen ditu eta funtsezko erronkak planteatzen ditu kazetaritzaren integritaterako eta osasun demokratikorako (Au, Ho & Chiu, 2021; Serrano-Puche, 2021). Fenomeno honi aurre egiteko, diziplina anitzeko ikuspegi bat behar da, kazetarien prestakuntza etikoa, publikoaren sentsibilizazioa eta informazio-arloan AAren jardutearen erregulazioa uztartuko dituenak. Neurri horien bidez bakarrik izango da posible hedabideen informazioarekiko konfiantza berrezartzea eta desinformazioak maila publikoan dituen ondorio disruptiboak arintzea.

3.4. Desinformazioaren aurka borrokatzeke baliabideak

Aro digitaleko desinformazioaren hazkundeak fenomeno horri aurre egiteko hainbat baliabide eta estrategia ezartzea bultzatu du. Horien artean hauek nabarmentzen dira: *fact-checking* deiturikoa, algoritmo aurreratuen erabilera (*random forests*, *bots* eta *chatbots*), monitorizazio-sistemak eta behatokiak, bai eta FANDANGO, Iberifier eta EDMO bezalako proiektu ezagunak ere. Gainera, alfabetatze mediatikoko programek eta kazetarien prestakuntzak garrantzia erdietsi dute eta, horri esker profesionalek modu autonomoan identifika ditzakete benetako albisteak gezurrezkoen aurrean,

haien gaitasun kritikoa eta etikoa indartuz etengabe aldatzen ari den ingurune mediatiko batean.

3.4.1 'Fact-checking'-aren garrantzia

Gailu mugikorren bidez informazio faltsua sortzeko erraztasunak *cheapfakes*-en fenomenoak ekarri du —modu amateurrean sortutako desinformazioak, alegia—, hezkuntza digitala eta alfabetatze mediatikoa hobetzeko beharra agerian uzten duena (Gamir-Ríos & Tarullo, 2022). Hezkuntza-tresna horien helburua herritarrak ahalduntzea da, egiazko informazioa eta manipulaturia bereizteko beharrezko gaitasunak emanez, eta eduki faltsuak ugaritzearen aurrean haien erresistentzia sendotuz.

Testuinguru horretan, *fact-checking* delakoa funtsezko praktika gisa sortu da eta honela definitzen da: «Argitaratutako edukietan akatsak, anbigüotasunak, gezurrak, zorrotasun falta edo zehaztasun ezak agerian uzteko datu-kazetaritzako teknikak aplikatzen dituen jarduera» (Ufarte-Ruiz, Peralta-García & Murcia-Verdú, 2018, 734. or.). Prozesu honi *debunking* ere esaten zaio eta baieztapenak zabaldu ondoren egiaztatzean datza, eduki faltsuak gutxieteko ahalegin hitzartua eginez (Lewandowsky, Cook & Lombardi, 2020).

Desinformazioaren aurkako borrokan *fact-checker*-rek duten papera oso ezaguna egin da azken hamarkadan (Hameleers & Van-der-Meer, 2020). Profesional espezializatu hauek zeregin erabakigarria betetzen dute integritate demokratikoa babesteko, ingurune digitalean dabilen informazio faltsua indargabetzen baitute (Moreno-Gil, Ramon-Vegas & Rodríguez-Martínez, 2021). Gainera, haien lana bereziki eraginkorra izan da politikarekin, klima-alaketarekin eta COVID-19aren pandemiarekin lotutako informazio faltsuak gezurtatzeari dagokionez (Salaverría et al., 2020; Pérez-Dasilva et al., 2020; Ramon-Vegas et al., 2020).

Hala eta guztiz ere, New York Times-ek (NYT) 2023ko irailean egindako txosten batek *fact-checking* ekimenetan geldialdi bat egon zitekeela adierazi zuen. Operazio horiek 2008ko 11tik 2022ko 424 arte igo baziren ere, kopurua pixka bat jaitsi da 417ra, eta horrek desinformazioaren aurkako borrokan ahaleginak indartu eta dibertsifikatu beharra azpimarratzen du.

3.4.2 Desinformazioaren aurka borrokatzeko proiektuak eta teknika berriak

Gaur egungo erronken aurrean, Europar Batasunak 2020tik finantzaturako Baliabide Digitalen Europako Behatokia (EDMO) bezalako proiektuek desinformazioaren aurkako borroka gidatu dute. EDMOk gertakari-egiaztatzaileekin, alfabetatze mediatikoaren adituekin eta ikertzaile akademikoekin lan egiten du, EBko 27 estatu kideak eta Norvegia hartzen

dituzten hamalau zentro nazional eta multinazionalen sare baten bidez (EDMO, 2023). Helburu nagusia tokiko desinformazio-kanpainak detektatzea eta aztertzea da, berariazko problematikei aurre egiteko ezagutza espezializatuak eskainiz.

Beste adibide nabarmen bat lberifier da, 2021etik 1.700 desinformazio baino gehiago egiaztatu dituen EDMOri lotutako proiektua. 2023ko ekaineko txosten baten arabera, desinformazio horien % 50 politikarekin lotuta dago, eta horietatik % 20 hauteskunde-prozesuetan zentratzen da. Ohiko beste gai batzuk klima-aldaketa, zibersegurtasuna eta migrazioak dira (Peña-Fernández, 2023). Ahalegin horrek agerian uzten du eskualdeko behatokiek arazo espezifikoak ikuspegi estrategiko batekin heltzeko duten gaitasuna.

Gainera, FANDANGO bezalako ekimenek Big Dataren erabilera oinarritutako sektore arteko ikuspegi bat proposatzen dute, hainbat iturritako datuak integratuz eduki faltsuak detektatu eta egiaztatzeko. Proiektu honek egiazko komunikazioa eratu nahi du hedabideen enpresak, gobernu-erakundeak eta gizarte-erakundeak aktiboki inplikaturik eta desinformazioaren aurrean lankidetzaz-are sendotuz.

Teknologiari dagokionez, algoritmo aurreratuak garatu dira albiste faltsuak detektatzeko, hala nola *random forests* eta neurona-sareak. Teknika horiek guztiek espero zen arrakasta lortu ez duten arren, ikerketek desinformazioa identifikatzeko orduan zehaztasuna eta abiadura hobetzeko ikuspegi berriak aztertzen jarraitzen dute (Pasquetto et al., 2022; Flores-Vivar, 2020).

AA erabiltzeak ere irauli egin ditu *fact-checking* ohiturak. Newtral-en ClaimHunter eta Maldita.es-en eta EFE Verifica-ren chatbot-en moduko tresnek desinformazio politiko eta sozialaren monitorizazioan eraginkorrak direla frogatu dute (Morrish, 2023). Ekimen horiek agerian utzi dute AA sortzaileak eduki faltsuen abiadurara eta konplexutasunera egokitzearen duen ahalmena, egiaztatzaileen erantzuteko gaitasuna hobetuz (Cybenko & Cybenko, 2018; Beckett, 2019).

3.4.3 Alfabetatze mediatikoa, informazio-alfabetatzea eta kazetarien prestakuntza

Alfabetatze mediatikoa funtsezko oinarria bihurtu da desinformazioari aurre egiteko. Ikuspegi honek norbanakoek eduki digitalarekin elkarrekaritzeko duten gaitasun kritikoa sendotzen du eta, gainera, informazio faltsuen aurkako borrokan erabilitako tresna teknologikoen osagarria ere bada (Amazeen & Bucy, 2019; Vraga, Bode & Tully, 2022). Europar Batasunak hainbat ekimenen bidez babestu ditu gaitasun horiek, hala nola *DigComp Framework for Digital Competence* (Ferrari, Punie & Brekko, 2013) eta Desinformazioaren aurkako Ekintza Plana (Europako Batzordea, 2018), era horretan, herritarren artean alfabetatzea sustatuz desinformazioa detektatzeko eta hari aurre egiteko.

Bestalde, UNESCOk *Media and Information Literate Citizens* hezkuntza-eredua sustatu du, hezitzaileen eta ikasleen gaitasun mediatikoak indartzeko, informazio manipulatuaren eragin negatiboak txikiagotzeko helburuarekin (Grizzle et al., 2021).

AAren inguruko etengabeko prestakuntzak eta haren ondorio etikoek ere garrantzia hartu dute. 2015ean sortutako *International Fact-Checking Network* (IFCN) erakundeak kazetariak egiaztapen-tresnen erabileran gaitzeko ekimenak gidatu ditu. Aldez aurretiko ikerketek nabarmentzen dute beharrezkoa dela ikasketa-planetan alfabetatze mediatikoko programak txertatzea eta azpimarratzen dute gaitasun kritikoak funtsezkoak direla informazio-ekosistema dinamiko batera egokitzeko (Celik, Muukkonen & Dogan, 2021; Jones-Jang et al., 2021).

Beraz, desinformazioaren aurkako hedabideek teknologia aurreratuak, alfabetatze-ekimenak eta sektoreen arteko lankidetzak uztartzen dituzte. Hala ere, eginahal honetan arrakasta lortzeko, beharrezkoa da AAeko gaitasunen etengabeko bilakaera eta konpromiso etikoa ingurune mediatikoaren erronka garaikideei aurre egiteko.

4. ONDORIOAK

Ikerketa honek AA eta kazetaritza bateratzetik sortzen den agertoki konplexua aztertzen du, eta, zehazki, teknologia emergente horiek hedatutako desinformazioaren problematika kritikoa jorratu du. Planteatutako galderei erantzutean, funtsezko ondorioak ateratzen dira eta paradigma mediatiko berri honetako erronkak eta aukerak nabarmentzen dituzte. AAK atzera bueltarik gabe eraldatu du informazioaren eremu profesionala eta aldekoen eta aurkakoen artean banatu ditu haren erabilerari buruzko iritziak. Erlazionatutako arriskuei buruzko adostasun profesionala dago eta desinformazioa da berehalako eta diziplina anitzeko erantzunak eskatzen dituen arazo larrigarrienetako bat.

Agerian geratu da estrategia integralen beharra. Kazetarientzako AAKo prestakuntza sendoa alfabetatze mediatiko eta informazio-alfabetatzeko programekin uztartzea funtsezko ikuspuntua da profesionalak aro digitalean behar diren trebetasunez hornitzeko. Ikuspuntu honek kazetariak ahaldundu ez ezik, audientziak indartu ere egiten ditu narratiba manipulatuak identifikatzeko eta horiei aurka egiteko gaitasunean.

Kazetaritzaren azterketa historikoak, aurrerapen teknologikoei dagokienez, funtsezko jardunbideen eraldaketa azpimarratzen du. Gatekeeper roletik gatewatcher rolara, AAK indar eragilea izan da albisteen ekoizpenean eta hedapenean, baina konplexutasun gehigarriak sartu ditu. Deepfake-ak, kazetaritza robota, oihartzun ganbera eta iragazki-burbuilak bezalako

fenomenoek areagotu egin dute informazioaren integritatea eta kazetaritzaren etika babesteko estrategia eraginkorrak garatzeko premia.

Albisteen ekoizpenean eta banaketan izandako joerek, bai eta AAre erabilerari lotutako dilema etikoek ere, erronka kritikoak mahaigaineratzen dituzte kazetaritzarekiko konfiantza publikoari dagokionez. Arestiko ikerketek nabarmendu dute desinformazioak, arrakala ideologikoak handitzeaz gain, demokrazien oinarriak ere ahultzen dituela, hedabideen sinesgarritasuna higatzen baitu. Beraz, dilema horiei arreta berezia eskaintzea funtsezkoa da hedabideekiko konfiantza eta informazio mediatikoaren sinesgarritasuna babesteko.

Sektore akademiko, profesional eta arautzaileen arteko lankidetzak aktibo baten bidez bakarrik eraiki liteke etorkizun bat, non kazetaritzak aurre egin eta gainditu ahal izango dituen AAK sustatutako desinformazioaren ondoriozko mehatxuak. Politika sendoak eta etikoak garatzea, desinformazioa egiaztatzeko eta detektatzeko teknologia aurreratuekin batera, urrats erabakigarria da ondorio kaltegarriak arintzeko.

Jasotako ebidentziek ikerketa honetan planteatutako hipotesiak babesten dituzte. Lehenik eta behin, AAK kazetaritzan izan duen eraginak desinformazioa areagotzen lagundu du, eta horrek erronka horri aurre egiteko komunikazio-profesionalak zorrotzago prestatzea eskatzen du. Bigarren, alfabetatze mediatiko eta informazio-alfabetatzeak, fact-checking-en moduko teknikak lagunduta, norbanakoak informazio faltsua identifikatzeko ahaldundutaz gain, desinformaziori testuinguru digital batean aurre egiteko irtenbide eraginkorra eta iraunkorra ere bihurtzen da.

Ekosistema mediatiko erresiliente bat eraikitzeko, ahalegin bateratuak egin behar dira erronka etikoak hartzeko, profesionalak gaitzeko eta kazetaritza digitalean konfiantza publikoa sustatzeko. Lido horretan, etengabeko egokitzapena, alfabetatze mediatikoa eta informazio-alfabetatzea eta aro digital honetako kazetaritza-praktiken bilakaera behar-beharrezkoak dira. Kazetaritza informatu, etiko eta erresistente batek soilik indargabetu dezake desinformazioa eta bermatu dezake etorkizunean informazio fidagarri baterako sarbidea.

8. ERREFERENTZIAK

- ADAMI, M. (2023). "Is ChatGPT a threat or an opportunity for journalism? Five AI experts weigh in". Reuters Institute, 23 March. <https://reutersinstitute.politics.ox.ac.uk/news/chatgpt-threat-or-opportunity-journalism-five-ai-experts-weigh>
- AGARWAL, U. (2018). "How Reuters is transforming journalism with artificial intelligence". *Analytics insight*, 11 March. <https://www.analyticsinsight.net/how-reuters-is-transforming-journalism-with-artificial-intelligence>

- ALJANABI, M.; GHAZI, M.; ALI, A.H.; ABED, S.A.; ChatGPT (2023). "ChatGpt: Open possibilities". *Iraqi journal for computer science and mathematics*, v. 4, n. 1, pp. 62-64. <https://doi.org/10.52866/ijcsm.2023.01.01.0018>
- ALKAISSI, H.; MCFARLANE, S. I. (2023). "Artificial hallucinations in ChatGPT: implications in scientific writing". *Cureus*, v. 15, n. 2. <https://doi.org/10.7759/cureus.35179>
- ALSHATER, M. M. (2022). "Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT". SSRN. <https://doi.org/10.2139/ssm.4312358>
- ALTMAN, S. (2022). "ChatGPT launched on Wednesday. Today it crossed 1 million users!". Twitter, December 5. [@sama]. <https://twitter.com/sama/status/1599668808285028353>
- AMAZEEN, M.A.; BUCY, E.P. (2019). "Conferring resistance to digital disinformation: the inoculating influence of procedural news knowledge". *Journal of broadcasting & electronic media*, v. 63, n. 3, pp. 415-432. <https://doi.org/10.1080/08838151.2019.1653101>
- ARIAS-JIMÉNEZ, B.; RODRÍGUEZ-HIDALGO, C.; MIER-SANMARTÍN, C.; CORONEL-SALAS, G. (2023). "Use of chatbots for news verification". In: López-López, Paulo-Carlos; Barredo, Daniel; Torres-Toukoumidis, Ángel; De-Santis, Andrea; Avilés, Óscar (eds.). *Communication and applied technologies. Smart innovation, systems and technologies*. Singapur: Springer Nature, v. 318, pp. 133-143. ISBN: 978 981 19 6347 6. https://doi.org/10.1007/978-981-19-6347-6_12
- AU, C-H.; HO, K.; CHIU, D. (2021). "The role of online misinformation and fake news in ideological polarization: barriers, catalysts, and implications". *Information systems frontiers*, v. 24, n. 2. <https://doi.org/10.1007/s10796-021-10133-9>
- AYDIN, Ö.; KARAARSLAN, E. (2022). "OpenAIChatGPT generated literature review: Digital twin in healthcare". In: Ay-din, Ömer (ed.). *Emerging computer technologies 2*. Izmir Akademi Dergisi, pp. 22-31. <https://doi.org/10.2139/ssm.4308687>
- BANG, Y.; CAHYAWIJAYA, S.; LEE, N.; DAI, W.; SU, D.; WILIE, B.; LOVENIA, H.; JI, Z.; YU, T.; CHUNG, W.; DO, Q. V.; XU, YA.; FUNG, P. (2023). "A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity". ArXiv preprint. ArXiv:2302.04023. <https://doi.org/10.48550/arXiv.2302.04023>
- BARBU, O. (2014). "Advertising, microtargeting and social media". *Procedia-social and behavioral sciences*, v. 163, pp. 44-49. <https://doi.org/10.1016/j.sbspro.2014.12.284>
- BECKETT, C. (2019). *New powers, new responsibilities: a global survey of journalism and artificial intelligence*. London School of Economics and Political Science. <https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities>
- BENNETT, W. L.; LIVINGSTON, S. (2018). "The disinformation order: disruptive communication and the decline of democratic institutions". *European journal of communication*, v. 33, n. 2, pp. 122-139. <https://doi.org/10.1177/0267323118760317>
- BRUNS, A. (2003). "Gatewatching, not gatekeeping: collaborative online news". *Media international Australia*, v. 107, n. 1, pp. 31-44. <https://doi.org/10.1177/1329878X0310700106>
- BODEN, M.A. (2022). *Inteligencia artificial*. Madrid: Turner. ISBN: 978 84 18895 35 7

- BOSHMAF, Y.; MUSKUKHOV, I.; BEZNOSOV, K.; RIPEANU, M. (2011). "The socialbot network: when bots socialize for fame and money". Twenty-seventh annual computer security applications conference [conference session], December 5th-9th. Orlando, USA <https://bit.ly/3YV2nGQ>
- BRADSHAW, S.; HOWARD, P.; KOLLANYI, B.; NEUDERT, L.M. (2020). "Sourcing and automation of political news and information over social media in the United States, 2016-2018". *Political communication*, v. 37, n. 2, pp. 173-193. <https://doi.org/10.1080/10584609.2019.1663322>
- BREWSTER, J.; ARVANITIS, L.; SADEGHI, M. (2023). "The next great misinformation superspreader: how ChatGPT could spread toxic misinformation at unprecedented scale. *NewsGuard*, 6 January. <https://www.newsguardtech.com/misinformation-monitor/jan-2023>
- BRUNDAGE, M.; AVIN, S.; CLARK, J.; TONER, H.; ECKERSLEY, P.; GARFINKEL, B.; DAFOE, A.; SCHARRE, P.; ZEITZOFF, T.; FILAR, B.; ANDERSON, H.; ROFF, H.; ALLEN, G.C.; STEINHARDT, J.; FLYNN, C.; HÉIGEARTAIGH, S.Ó.; BEARD, S.; BELFIELD, H.; FARQUHAR, S.; LYLE, C.; CROOTOFF, R.; EVANS, O.; PAGE, M.; BRYSON, J.; YAMPOLSKIY, R.; AMODEI, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf>
- CANAVILHAS, J. (2022). "Inteligencia artificial aplicada al periodismo: traducción automática y recomendación de contenidos en el proyecto 'A European Perspective' (UER)". *Revista latina de comunicación social*, n. 80. <https://www.doi.org/10.4185/RLCS-2022-1534>
- CARLSON, M.; LEWIS, S.C. (eds.). (2015). *Boundaries of journalism: professionalism, practices and participation*. London: Routledge. ISBN: 978 1 138020 67 2
- CASERO-RIPOLLÉS, A.; DOMÉNECH-FABREGAT, H.; ALONSO-MUÑOZ, L. (2023). "Percepciones de la ciudadanía española ante la desinformación en tiempos de la Covid-19: efectos y mecanismos de lucha contra las noticias falsas". *Icono 14*, v. 21, n. 1. <https://doi.org/10.7195/ri14.v21i1.1988>
- CELIK, I., MUUKKONEN, H., & DOGAN, S. (2021). A model for understanding new media literacy: Epistemological beliefs and social media use. *Library & Information Science Research*, 43(4), 101125.
- CHADWICK, A. (2013). *The hybrid media system: politics and power*. Oxford: Oxford University Press. ISBN: 978 0 190696 73 3
- CHADWICK, A.; VACCARI, C.; O'LOUGHLIN, B. (2018). "Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing". *New media & society*, v. 20, n. 11, pp. 4255-4274. <https://doi.org/10.1177/1461444818769689>
- CINELLI, M.; DE-FRANCISCI-MORALES, G.; GALEAZZI, A.; QUATTROCIOCCHI, W.; STARNINI, M. (2021). "The echo chamber effect on social media". *Proceedings of the National Academy of Sciences*, v. 118, n. 9 <https://doi.org/10.1073/pnas.2023301118>
- CLAYTON, K., BLAIR, S., BUSAM, J. A., FORSTNER, S., GLANCE, J., GREEN, G., NYHAN, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political behavior*, 42, 1073-1095.

- CLOUTIER, J. (1973). *La communication audio-scripto-visuelle à l'heure des self média*. Montreal: Les Presses de l'Université de Montreal.
- COHEN, S.; HAMILTON, J.T.; TURNER, F. (2011). "Computational journalism". *Communications of the ACM*, v. 54, n. 10, pp. 66-71. <https://doi.org/10.1145/2001269.2001288>
- COIRO, J., KNOBEL, M., LANKSHEAR, C., LEU, D.J. (Eds.). (2014). *Handbook of research on new literacies*. Routledge.
- COLEMAN, S. (2012). "Believing the news: From sinking trust to atrophied efficacy". *European journal of communication*, v. 27, n. 1, pp. 35-45 <https://doi.org/10.1177/0267323112438806>
- COLLEY, T.P.; GRANELLI, F.; ALTHUIS, J.(2020). "Disinformation's societal impact: Britain, Covid and beyond". *Defence strategic communications*, v. 8, n. 1, pp. 89-140.<https://doi.org/10.1080/10.30966/2018.RIGA.8.3>
- COMISIÓN EUROPEA (2018a). "La lucha contra la desinformación en línea: un enfoque europeo". Comunicación de la Comisión al Parlamento europeo, al Comité económico y social europeo y al Comité de las regiones. COM(2018)236.[https://ec.europa.eu/transparency/documents-register/detail?ref=COM\(2018\)236](https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2018)236)
- COMISIÓN EUROPEA (2018b). "Action plan against disinformation". Join (2018) 36 final.https://eeas.europa.eu/sites/eeas/files/action_plan_against_disinformation.pdf
- COMISIÓN EUROPEA (2018c). "Code of practice on disinformation". <https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation>
- COMISIÓN EUROPEA (2022). Strengthened code of practice disinformation. <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>
- COSTA-SÁNCHEZ, C.; LÓPEZ-GARCÍA, X. (2020). "Comunicación y crisis del coronavirus en España: primeras lecciones". *Profesional de la información*, v. 29, n. 3, e290304 <https://doi.org/10.3145/epi.2020.may.04>
- CRAWFORD, K. (2021). *The atlas of AI: power, politics, and the planetary costs of artificial intelligence*. Yale University Press. ISBN: 978 0 300252392 <https://doi.org/10.2307/j.ctv1ghv45t>
- CYBENKO, A.K.; CYBENKO, G. (2018). "AI and fake news". *IEEE intelligent systems*, n. 33. <https://doi.org/10.1109/MIS.2018.2877280>
- DEUZE, M.; WITSCHGE, T. (2018). "Beyond journalism: Theorizing the transformation of journalism". *Journalism*, v. 19, n. 2, pp. 165-181. <https://doi.org/10.1177/1464884916688550>
- EDELMAN (2023). Edelman Trust Barometer. Navigating a polarized world <https://www.edelman.com/trust/2023/trust-barometer>
- EDMO (2023). European Digital Media Observatory. [Disponible en línea](<https://edmo.eu>)
- ENGUIX-OLIVER, SALVADOR (2020). "Efectos de la digitalización en la relación entre periodistas y políticos: desintermediación y pactos comunicativos". *Pasajes*, v. 59, pp. 19-30. <https://roderic.uv.es/bitstream/handle/10550/76566/7550103.pdf>

- FAWZI, N.; STEINDL, N.; OBERMAIER, M.A; PROCHAZKA, F.; ARLT, D.; BLÖBAUM, B.; DOHLE, M.; ENGELKE, K. M.; HANITZSCH, T.; JACKOB, N.; JAKOBS, I.; KLAWIER, T.; POST, S.; REINEMANN, C.; SCHWEIGER, W.; ZIEGELE, M. (2021). "Concepts, causes and consequences of trust in news media - a literature review and framework". *Annals of the international communication association*, v. 45, n. 2, pp. 154-174. <https://doi.org/10.1080/23808985.2021.1960181>
- FERNÁNDEZ-BARRERO, Á.; LÓPEZ-REDONDO, I. (2022). "Verification processes in fake news era. Some examples about Covid-19". *Ambitos. Revista internacional de comunicación*, n. 57, pp. 124-137. <https://doi.org/10.12795/Ambitos.2022.i57.07>
- FERRARI, A.; PUNIE, Y.; BREKKO, B.N. (2013). DigComp: A framework for developing and understanding digital competence in Europe <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC83167/lb-na-26035-enn.pdf>
- FLETCHER, R.; PARK, S. (2017). "The impact of trust in the news media on online news consumption and participation". *Digital journalism*, v. 5, n. 10, pp. 1281-1299. <https://doi.org/10.1080/21670811.2017.1279979>
- FLORES-VIVAR, J.M. (2019). "Inteligencia artificial y periodismo: diluyendo el impacto de la desinformación y las noticias falsas a través de los bots". *Doxa comunicación*, n. 29, pp. 197-212. <https://doi.org/10.31921/doxacom.n29a10>
- FLORES-VIVAR, J.M.(2020). "Datos masivos, algoritmización y nuevos medios frente a desinformación y fake news. Bots para minimizar el impacto en las organizaciones". *Comunicación y hombre*, n. 16, pp. 101-114. <https://doi.org/10.32466/eufv-cyh.2020.16.601.101-114>
- GAMIR-RÍOS, J.; TARULLO, R. (2022). "Predominio de las cheapfakes en redes sociales: Complejidad técnica y funciones textuales de la desinformación desmentida en Argentina durante 2020". *adComunica*, n. 23, pp. 97-118 <https://doi.org/10.6035/adcomunic%EF%BB%BFa%EF%BB%BF.6299>
- GARCÍA-AVILÉS, J.A. (2019). "El impacto de la inteligencia artificial en el periodismo". Máster en innovación y periodismo UMH blog, 2 noviembre. <http://mip.umh.es/blog/2019/11/02/el-impacto-de-la-inteligencia-artificial-en-el-periodismo>
- GARCÍA-GALERA, M.C.; DEL-HOYO-HURTADO, M.; BLANCO-ALFONSO, I.(2020). "Disinformation and communicative intent: a proposal for fake news classification in professional journalistic environments". *Revista mediterranea de comunicacion*, v. 11, n. 2, pp. 105-118. <https://doi.org/10.14198/MEDCOM2020.11.2.16>
- GARCÍA-MARÍN, D.; SALVAT-MARTINREY, G. (2022). "Tendencias en la producción científica sobre desinformación en España. Revisión sistematizada de la literatura (2016-2021)". *Adcomunica*, n. 23, pp. 23-50. <https://doi.org/10.6035/adcomunica.6045>
- GAZIANO, C.; MCGRATH, K. (1986). "Measuring the concept of credibility". *Journalism and mass communication quarterly*, v. 63, n. 3, pp. 451-462. <https://doi.org/10.1177/107769908606300301>
- GILAT, R.; COLE, B. J. (2023). "How will artificial intelligence affect scientific writing, reviewing and editing? The future is here". *Arthroscopy*, v. 39, n. 5, pp. 1119-1120. <https://doi.org/10.1016/j.arthro.2023.01.014>
- GILLESPIE, T. (2010). "The politics of 'platforms'". *New media & society*, v. 12, n. 3, pp.347-364. <https://doi.org/10.1177/1461444809342738>

- GOLAN, R.; REDDY, R.; MUTHIGI, A.; RAMASAMY, R. (2023). "Artificial intelligence in academic writing: a paradigm-shifting technological advance". *Nature reviews urology*, 24 February. <https://doi.org/10.1038/s41585-023-00746-x>
- GONÇALVES, A.; MELO, P.V. (2022). "Artificial intelligence and journalism: An approach to the Portuguese context". *Fonseca, journal of communication*, n. 25, pp. 23-24 <https://doi.org/10.14201/fjc.29682>
- GOOGLE(2023). Bard. <https://bard.google.com>
- GRAU, M. (2020). "New WhatsApp chatbot unleashes power of worldwide fact-checking organizations to fight COVID-19 misinformation on the platform". International Fact-Checking Network (IFCN), Poynter Institute, 4 May. <https://www.poynter.org/fact-checking/2020/poynters-international-fact-checking-network-launches-whatsapp-chatbot-to-fight-covid-19-misinformation-leveraging-database-of-more-than-4000-hoaxes>
- GUZMAN, A.L.; LEWIS, S.C. (2020). "Artificial intelligence and communication: a human-machine communication research agenda". *New media & society*, v. 22, n. 1, pp. 70-86. <https://doi.org/10.1177/1461444819858691>
- HAMELEERS, M.; VAN-DER-MEER, T. G. L. A. (2020). "Misinformation and polarization in a high-choice media environment: how effective are political fact-checkers?" *Communication research*, v. 47, n. 2, pp. 227-250. <https://doi.org/10.1177/0093650218819671>
- HANITZSCH, T.; VAN-DALEN, A.; STEINDL, N. (2018). "Caught in the nexus: A comparative and longitudinal analysis of public trust in the press". *The international journal of press/politics*, v. 23, n. 1, pp. 3-23. <https://doi.org/10.1177/1940161217740695>
- HANSEN, M.; ROCA-SALES, M.; KEEGAN, J.; KING, G. (2017). Artificial intelligence: Practice and implications for journalism. Columbia Journalism School. <https://doi.org/10.7916/D8X92PRD>
- HAZARD-OWEN, L. (2023). "Google is changing up search. What does that mean for news publishers?". *NiemanLab*, 11 May. <https://www.niemanlab.org/2023/05/google-is-changing-up-search-what-does-that-mean-for-news-publishers>
- HIMMA-KADAKAS, M.; OJAMETS, I. (2022). "Debunking false information: investigating journalists' fact-checking skills". *Digital journalism*, v. 10, n. 5, pp. 866-887. <https://doi.org/10.1080/21670811.2022.2043173>
- HOBBS, R. (2017). *Create to learn: Introduction to digital literacy*. John Wiley & Sons.
- HOVLAND, C.I.; WEISS, W.(1951). "The influence of source credibility on communication effectiveness". *The public opinion quarterly*, v. 15, n. 4, pp. 635-650. <http://www.jstor.org/stable/2745952>
- HUANG, J.; TAN, M. (2023). "The role of ChatGPT in scientific communication: writing better scientific review articles". *American journal of cancer research*, v. 13, n. 4, pp. 1148-1154. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10164801>
- JANKOWSKI, N. W. (2018). "Researching fake news: a selective examination of empirical studies". *Javnost-The public*, v. 25, n. 1-2, pp. 248-255. <https://doi.org/10.1080/13183222.2018.1418964>
- JENKINS, H., C., K., PURUSHOTMA, R., ROBISON, A.J. & WEIGEL, M. (2006). *Confronting the Challenges of Participatory Culture: Media Education For the*

- 21st Century. Chicago: The MacArthur Foundation. Retrieved July 10, 2007 from <http://www.newmedialiteracies.org/files/working/NMLWhitePaper.pdf>
- JONES-JANG, S. M., MORTENSEN, T., & LIU, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American behavioral scientist*, 65(2), 371-388.
- KAPANTAI, E.; CHRISTOPOULOU, A.; BERBERIDIS, C.; PERISTERAS, V. (2021). "A systematic literature review on disinformation: toward a unified taxonomical framework". *New media & society*, v. 23, n. 5, pp. 1301-1326. <https://doi.org/10.1177/1461444820959296>
- KARLSEN, J.; STAVELIN, E.(2013). "Computational journalism in Norwegian newsrooms". *Journalism practice*, v. 8, n. 1, pp. 34-48 <https://doi.org/10.1080/17512786.2013.813190>
- KIM B., XIONG A., LEE D., HAN K. (2021) A systematic review onfakenews research through the lens of news creation and consumption: Research efforts, challenges, and future directions. *PLoS ONE*16(12): e0260080. <https://doi.org/10.1371/journal.pone.0260080>
- KIM, D.; KIM, S. (2017). "Newspaper companies' determinants in adopting robot journalism". *Technological forecasting and social change*, v. 117, pp. 184-195. <https://doi.org/10.1016/j.techfore.2016.12.002>
- KIOUSIS, S. (2001). "Public trust or mistrust? Perceptions of media credibility in the information age". *Mass communication & society*,v. 4, n. 4, pp. 381-403. https://doi.org/10.1207/S15327825MCS0404_4
- KOHRING, M.; MATTHES, J. (2007). "Trust in news media: development and validation of a multidimensional scale". *Communication research*, v. 34, n. 2, pp. 231-252. <https://doi.org/10.1177/0093650206298071>
- LEÓN, B.; LÓPEZ-GOÑI, I.; SALAVERRÍA, R. (2022). "The Covid-19 catastrophe: a science communication mess?". *Church, communication and culture*, v. 7, n. 1, pp. 6-22. <https://doi.org/10.1080/23753234.2022.2031236>
- LEWANDOWSKY, S., COOK, J., ECKER, U. K., LEWANDOWSKY, S., COOK, J., ECKER, U. K. H., & NEWMAN, E. J. (2020). Under the Hood of The Debunking Handbook 2020: A consensus-based handbook of recommendations for correcting or preventing misinformation. Center for Climate Change Communication. <https://www.climatechangecommunication.org/wp-content/uploads/2020/10/DB2020paper.Pdf>
- LEWANDOWSKY, S.; ECKER, U.K.; SEIFERT, C.M.; SCHWARZ, N.; COOK, J.(2012). "Misinformation and its correction: continued influence and successful debiasing". *Psychological science in the public interest*, v. 13, n. 3, pp. 106-131 <https://doi.org/10.1177/1529100612451018>
- LICHTERMAN, J. (2017). "Want to bring automation to your newsroom? A new AP report details best practices". *NiemanLab*, 5 April. <https://www.niemanlab.org/2017/04/want-to-bring-automation-to-your-newsroom-a-new-ap-report-details-best-practices>
- LOCK, S.(2022). "What is ai chatbot phenomenon ChatGPT and could it replace humans?". *The Guardian*, December 5. <https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans>

- LOO-VÁZQUEZ, J.R.; GÁMEZ-PAZ, D.; LAMARQUE-VEGA, E.; DOMÍNGUEZ-QUEVEDO, L.; HABER-GUE-RRRA, Y. (2016). "Del gatekeeper al content curator: cambiar algo para que no cambie nada". *Comunicología cubana*, v. 20, n. 92, pp. 1149-1171. <https://www.revistarazonypalabra.org/index.php/ryp/article/view/293>
- LÓPEZ-JIMÉNEZ, E.A.; OUARIACHI, T. (2020). "An exploration of the impact of artificial intelligence (AI) and automation for communication professionals". *Journal of information, communication and ethics in society*, v. 19, n. 2, pp. 249-267. <https://doi.org/10.1108/JICES-03-2020-0034>
- LÓPEZ-MARTÍN, Á.; GÓMEZ-CALDERÓN, B.; CÓRDOBA-CABÚS, A. (2023). "La desinformación en auge: un análisis de los bulos sobre política española". Visual review. *International visual culture review*, v. 10, n. 3. <https://doi.org/10.37467/revvisual.v10.4596>
- LÓPEZ-PAN, F.; RODRÍGUEZ-RODRÍGUEZ, J.M. (2020). "El fact checking en España. Plataformas, prácticas y rasgos distintivos". *Estudios sobre el mensaje periodístico*, v. 26, n. 3, pp. 1045-1065. <https://doi.org/10.5209/ESMP65246>
- MAGALLÓN-ROSA, R., FERNÁNDEZ-CASTRILLO, C., GARRIGA, M. (2023). Fact-checking in war: Types of hoaxes and trends from a year of disinformation in the Russo-Ukrainian war. *Profesional de la información/Information Professional*, 32(5).
- MANFREDI-SÁNCHEZ, J. LUIS; UFARTE-RUIZ, M.J. (2020). "Inteligencia artificial y periodismo: una herramienta con-tra la desinformación". *Revista Cidob d'afers internacionals*, n. 124, pp. 49-72. <https://doi.org/10.24241/rcai.2020.124.1.49>
- MARWICK, A. E., LEWIS, R. (2017). *Media manipulation and disinformation online*.
- MASIP, P.; SUAUI, J.; RUIZ-CABALLERO, C. (2020). "Percepciones sobre medios de comunicación y desinformación: ideología y polarización en el sistema mediático español". *Profesional de la información*, v. 29, n. 5, e290527. <https://doi.org/10.3145/epi.2020.sep.27>
- MEDIA AND INFORMATION LITERATE CITIZENS: THINK CRITICALLY, CLICK WISELY! (s. f.). The Communication Initiative Network. <https://www.comminet.com/global/content/media-and-information-literate-citizens-think-critically-click-wisely>
- METZGER, M. J., FLANAGIN A. J., ZWARUN L. . (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers and Education* 41:271-90
- MEYER, P. (1988). "Defining and measuring credibility of newspapers: developing an index". *Journalism & mass communication quarterly*, v. 65, n. 3, pp. 567-574 <https://doi.org/10.1177/107769908806500301>
- MÍGUEZ-GONZÁLEZ, M.I.; ABUÍN-PENAS, J.; PÉREZ-SEOANE, J. (2021). "¿Cómo utilizan los fact-checkers las redes sociales para combatir la desinformación? Análisis de la actividad de los fact-checkers iberoamericanos en Insta-gram". En: Blanco-Pérez, Manuel (ed.). *El progreso de la comunicación en la era de los prosumidores*. Madrid: Dykinson SL, pp. 15-39. ISBN: 978 84 1377 644 6 <https://www.dykinson.com/libros/el-progreso-de-la-comunicacion-en-la-era-de-los-prosumidores/9788413776446>
- MONTEMAYOR-RODRÍGUEZ, N.; GARCÍA-JIMÉNEZ, A. (2021). "Percepción de los periodistas sobre la desinformación y las rutinas profesionales en la era digital". *Revista general de información y documentación*, v. 31, n. 2, pp. 601-619. <https://doi.org/10.5209/rgid.79460>

- MOOSHAMMER, S. (2022). "There are (almost) no robots in journalism. An attempt at a differentiated classification and terminology of automation in journalism on the base of the concept of distributed and gradualised action". *Publizistik*, v. 67, pp. 487-515 <https://doi.org/10.1007/s11616-022-00757-5>
- MORENO-GIL, V.; RAMON-VEGAS, X.; RODRÍGUEZ-MARTÍNEZ, R. (2021). "Fact-checking interventions as counter-offensives to disinformation growth: Standards, values, and practices in Latin America and Spain". *Media and communication*, v. 9, n. 1, pp. 251-263. <https://doi.org/10.17645/mac.v9i1.3443>
- MORRISH, L.(2023). Fact-checkers are scrambling to fight disinformation with AI. *The wired*, 1 February. <https://www.wired.co.uk/article/fact-checkers-ai-chatgpt-misinformation>
- MURTHY, D. POWELL, A.B.; TINATI, R.; ANSTEAD, N.; CARR, L.; HALFORD, S. J.; WEAL, M. (2016). "Automation, algorithms, and politics. Bots and political influence: A sociotechnical investigation of social network capital". *IJOC*, n. 10, v. 2016, pp. 4952-4971. <https://ijoc.org/index.php/ijoc/article/view/6271>
- NEWMAN, N.; FLETCHER, R.; ROBERTSON, C. T.; EDDY, K.; NIELSEN, R.K. (2022). Digital news report 2022. <https://reutersinstitute.politics.ox.ac.uk/es/digital-news-report/2022>
- NOAIN-SÁNCHEZ, A. (2022). "Addressing the impact of artificial intelligence on journalism: the perception of experts, journalists and academics". *Communication & society*, v. 35, n. 3, pp. 105-21. <https://doi.org/10.15581/003.35.3.105-121>
- ORBEGOZO-TERRADILLOS, J.; MORALES-I-GRAS, J.; LARRONDO-URETA, A.(2020). "Desinformación en redes sociales: ¿compartimentos estancos o espacios dialécticos? El caso Luther King, Quim Torra y El confidencial". *Revista mediterránea de comunicación*, v. 11, n. 2, pp. 55-69 <https://doi.org/10.14198/MEDCOM2020.11.2.2>
- ORGANIZACIÓN MUNDIAL DE LA SALUD (2020). Gestión de la infodemia sobre la Covid-19: Promover comportamientos saludables y mitigar los daños derivados de la información incorrecta y falsa. <https://bit.ly/3NDLQ44>
- OTTO, K.; KÖHLER, A.(eds.) (2018). Trust in media and journalism: Empirical perspectives on ethics, norms, impacts and populism in Europe. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-20765-6>
- PALOMO, B.; SEDANO-AMUNDARAIN, J.A. (2018). "WhatsApp como herramienta de verificación de fake news. El caso de B de Bulo". *Revista latina de comunicación social*, n. 73, pp. 1384-1397. <https://doi.org/10.4185/RLCS-2018-1312>
- PARISER, E. (2011). The filter bubble: what the Internet is hiding from you. Penguin. ISBN: 978 1 59420300 8
- PASQUETTO, I. V.; JAHANI, E.; ATREJA, S.; BAUM, M. (2022). "Social debunking of misinformation on WhatsApp: the case for strong and in-group ties". *Proceedings of the ACM on human-computer interaction*, v. 6. <https://doi.org/10.1145/3512964>
- PEÑA-FERNÁNDEZ, S.; MESO-AYERDI, K.; LARRONDO-URENA, A.; DÍAZ-NOCI, J. (2023). "Sin periodistas, no hay periodismo. La dimensión social de la inteligencia artificial generativa en los medios de comunicación". *Profesional de la información*, v. 32, n. 2, e320227 <https://doi.org/10.3145/epi.2023.mar.27>

- PÉREZ-DASILVA, J.A.; MESO-AYERDI, K.; MENDIGUREN-GALDOSPÍN, T. (2020). "Fake news and coronavi-rus: Detecting key players and trends through analysis of Twitter conversations". *Profesional de la informacion*, v. 29, n. 3. <https://doi.org/10.3145/epi.2020.may.08>
- PÉREZ-ESCOLAR, M.; LILLEKER, D.; TAPIA-FRADE, A. (2023). "A systematic literature review of the phenome-non of disinformation and misinformation". *Media and communication*, v. 11, n. 2, pp. 76-87. <https://doi.org/10.17645/mac.v11i2.6453>
- PÉREZ-LATRE, F.J. (2022). *Crisis de confianza (2007-2022). El descrédito de los medios*. Pamplona: Eunsa. ISBN: 978 84 313 3792 6
- PIHLAJARINNE, T.; ALÉN-SAVIKKO, A. (2022). *Artificial intelligence and the media. Reconsidering rights and responsibilities*. Cheltenham, UK: Edward Elgar Publishing. <https://www.elgaronline.com/view/edcoll/9781839109966/9781839109966.xml>
- PIRA, F. (2019). "Las elecciones europeas de 2019 y las fake news". *Barataria*, n. 26, pp. 67-83. <https://doi.org/10.20932/barataria.v0i26.510>
- POSETTI, J.; MATTHEWS, A. (2018). A short guide to the history of 'fake news' and disinformation. International Center for Journalists, July 2018 https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf
- PREGO, C. (2023). "Elon Musk lanza X.AI, su propia compañía de IA con un objetivo claro: competir con OpenAI". *Xataka*, 15 abril. <https://www.xataka.com/empresas-y-economia/elon-musk-lanza-su-propia-compania-inteligencia-artificial-x-ai-que-sabemos-su-plan-para-competir-openai>
- RAMON-VEGAS, X.; MAURI-RÍOS, M.; RODRÍGUEZ-MARTÍNEZ, R.(2020). "Redes sociales y plataformas de fact-checking contra la desinformación sobre la COVID-19". *Hipertext.net*, v.21, pp. 79-92. <https://doi.org/10.31009/hipertext.net.2020.i21.07>
- RODRÍGUEZ, M.; GIRI, L. (2021). "Desafíos teóricos cruciales para la comunicación pública de la ciencia y la tecnología post pandemia en Iberoamérica". *Revista iberoamericana de ciencia, tecnología y sociedad*, v. 16, pp. 25-39. <http://www.revistacts.net/contenido/numero-numero-especial-2021/desafios-teoricos-cruciales-para-la-comunicacion-publica-de-la-ciencia-y-la-tecnologia-post-pandemia-en-iberoamerica>
- RODRÍGUEZ-PÉREZ, C. (2020). "Una reflexión sobre la epistemología del fact-checking journalism: retos y dilemas". *Revista de comunicación*, v. 19, n. 1, pp. 243-258. <https://doi.org/10.26441/RC19.1-2020-A14>
- ROJAS-TORRIJOS, J.L.; TOURAL-BRAN, C. (2019). "Periodismo deportivo automatizado. Estudio de caso de AnaFut, el bot desarrollado por El Confidencial para la escritura de crónicas de fútbol". *Doxa comunicación.Revista interdisciplinar de estudios de comunicación y ciencias sociales*, n. 29, pp. 235-254. <https://doi.org/10.31921/doxacom.n29a12>
- ROOSE, KEVIN (2022). "The brilliance and weirdness of ChatGPT". *The New York Times*. <https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html>
- ROSSETTI, M.; ZAMAN, T. (2023). "Bots, disinformation, and the first impeachment of U.S. President Donald Trump". *Plos one*, v. 18, n. 5. <https://doi.org/10.1371/journal.pone.0283971>

- RUSSELL, S.; NORVIG, P. (2022). *Artificial intelligence: a modern approach*. Fourth ed. London: Pearson Education. ISBN: 978 1 292 40117 1
- SÁDABA, C.; SALAVERRÍA, R. (2023). "Combatir la desinformación con alfabetización mediática: análisis de las tendencias en la Unión Europea". *Revista latina de comunicación social*, n. 81, pp. 17-33. <https://doi.org/10.4185/rlds-2023-1552>
- SALAVERRÍA, R.; BUSLÓN, N.; LÓPEZ-PAN, F.; LEÓN, B.; LÓPEZ-GOÑI, I.; ERVITI, M.C. (2020). "Desinformación en tiempos de pandemia: Tipología de los bulos sobre la Covid-19". *Profesional de la información*, v. 29, n. 3, e290315. <https://doi.org/10.3145/epi.2020.may.15>
- SÁNCHEZ-GARCÍA, P.; MERAYO-ÁLVAREZ, N.; CALVO-BARBERO, C.; DIEZ-GRACIA, A. (2023). "Spanish technological development of artificial intelligence applied to journalism: companies and tools for documentation, production and distribution of information". *Profesional de la información*, v. 32, n. 2, e320208 <https://doi.org/10.3145/epi.2023.mar.08>
- SÁNCHEZ-GONZALES, H.M.; SÁNCHEZ-GONZÁLEZ, M. (2017). "Los bots como servicio de noticias y de conectividad emocional con las audiencias. El caso de Politibot". *Doxa comunicación. Revista interdisciplinar de estudios de comunicación y ciencias sociales*, v. 25, pp. 63-84. <https://doi.org/10.31921/doxacom.n25a3>
- SERRANO-PUCHE, J. (2017). "Credibilidad y confianza en los medios de comunicación: un panorama de la situación en España". En: González, Marta; Valderrama, Mónica (eds.). *Discursos comunicativos persuasivos hoy*, pp. 427-438. Madrid: Tecnos. ISBN: 978 9978 375 65 5
- SMUHA, N. A. (2021). From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence. *Law, Innovation and Technology*, 13(1), 57-84.
- STARBIRD, K.; ARIF, A.; WILSON, T. (2019). "Disinformation as collaborative work: surfacing the participatory nature of strategic information operations". *Proceedings of the ACM on human-computer interaction*, v. 3, n. 127. <https://doi.org/10.1145/3359229>
- STRÖMBÄCK, J.; TSFATI, Y.; BOOMGAARDEN, H.; DAMSTRA, A.; LINDGREN, E.; VLIEGENTHART, R.; LINDHOLM, T. (2020). "News media trust and its impact on media use: toward a framework for future research". *Annals of the International Communication Association*, v. 44, n. 2, pp. 139-156. <https://doi.org/10.1080/23808985.2020.1755338>
- SUBIRES-MANCERA, M.P. (2017). "La lucha contra las noticias falsas en Internet". En: Larrondo-Ureta, Aina-ra; Meso-Ayerdi, Koldobika; Peña-Fernández, Simón (coord.). *IX Congreso Internacional de Ciberperiodismo: Innovación y emprendimiento al servicio de las audiencias*. Bilbao: Universidad del País Vasco, pp. 512-526. ISBN : 978 84 9082 783 3 <https://web-argitalpena.adm.ehu.es/listaproductos.asp?IdProducts=USPDF177833>
- SUNSTEIN, C. R. (2017). Nudges that fail. *Behavioural public policy*, 1(1), 4-25.
- TEJEDOR, S.; VILA, P. (2021). "Exo journalism: A Conceptual approach to a hybrid formula between journalism and artificial intelligence". *Journalism and media*, v. 2, n. 4, pp. 830-840 <https://doi.org/10.3390/journalmedia2040048>
- THURMAN, N.; DÖRR, K., & KUNERT, J. (2017). When reporters get hands-on with Robo-Writing. *Digital Journalism*, 5(10), 1240-1259. <https://doi.org/10.1080/21670811.2017.1289819>

- TSFATI, Y.; ARIELY, G. (2014). "Individual and contextual correlates of trust in media across 44 countries". *Communication research*, v. 41, n. 6, pp. 760-782. <https://doi.org/10.1177/0093650213485972>
- UFARTE-RUIZ, M.J.; FIEIRAS-CEIDE, C.; TÚÑEZ-LÓPEZ, M. (2020). "La enseñanza-aprendizaje del periodismo automatizado en instituciones públicas: estudios, propuestas de viabilidad y perspectivas de impacto de la IA". *Análisi*, n. 62, pp. 131-146. <https://doi.org/10.5565/rev/analisi.3289>
- UFARTE-RUIZ, M.J.; PERALTA-GARCÍA, L.; MURCIA-VERDÚ, F.J. (2018). "Fact checking: un nuevo desafío del periodismo". *El profesional de la información*, v. 27, n. 4, pp. 733-741. <https://doi.org/10.3145/epi.2018.jul.02>
- UFARTE-RUIZ, M.J.; MANFREDI-SÁNCHEZ, J.L. (2019). "Algoritmos y bots aplicados al periodismo. El caso de Narrativa Inteligencia Artificial: estructura, producción y calidad informativa". *Doxa comunicación. Revista interdisciplinar de estudios de comunicación y ciencias sociales*, v. 29, pp. 213-233 <https://doi.org/10.31921/doxacom.n29a11>
- UFARTE-RUIZ, M.J.; MURCIA-VERDÚ, F.J.; TÚÑEZ-LÓPEZ, J.M. (2023). "Use of artificial intelligence in synthetic media: first newsrooms without journalists". *Profesional de la información*, v. 32, n. 2 <https://doi.org/10.3145/epi.2023.mar.03>
- VÁLLEZ, M.; CODINA, LL. (2018). "Periodismo computacional: evolución, casos y herramientas". *Profesional de la información*, v. 27, n. 4, pp. 759-768 <https://doi.org/10.3145/epi.2018.jul.05>
- VAN-AELST, P.; STRÖMBÄCK, J.R.; AALBERG, T.; ESSER, F. (2017). "Political communication in a high-choice media environment: a challenge for democracy?". *Annals of the international communication association*, v. 41, n. 1, pp. 3-27. <https://doi.org/10.1080/23808985.2017.1288551>
- VAN-NOORDT, C.; MISURACA, G. (2022). "Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union". *Government information quarterly*, v. 39, n. 3, e101714 <https://doi.org/10.1016/j.giq.2022.101714>
- VOSOUGHI, S.; ROY, D.; ARAL, S. (2018). "The spread of true and false news online". *Science*, v. 359, n. 6380, pp. 1146-1151 <https://doi.org/10.1126/science.aap9559>
- VRAGA, E.K.; BODE, L.; TULLY, M. (2022). "Creating news literacy messages to enhance expert corrections of misinformation on Twitter". *Communication research*, v. 49, n. 2, pp. 245-267. <https://doi.org/10.1177/0093650219898094>
- WAISBORD, S. (2018). "Truth is what happens to news: on journalism, fake news, and post-truth". *Journalism studies*, v. 19, n. 13, pp. 1866-1878 <https://doi.org/10.1080/1461670X.2018.1492881>
- WALTER, N.; COHEN, J.; HOLBERT, R. L.; MORAG, Y. (2020). "Fact-checking: a meta-analysis of what works and for whom". *Political communication*, v. 37, n. 3, pp. 350-375. <https://doi.org/10.1080/10584609.2019.1668894>
- WARDLE, C.; DERAKHSHAN, H. (2017). Information disorder: toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe. <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>

Sarrionandia, B: Kazetarientzako tresnak adimen artifizialak indartutako desinformazioaren...

- WU, J. (2016). The state of media studies in China. *Television & New Media*, 17(5), 392-396. <https://doi.org/10.1177/1527476416630111>
- WU, S., TANDOC, E. C., SALMON, C. T. (2019). When Journalism and automation intersect: Assessing the influence of the technological field on contemporary newsrooms. *Journalism Practice*, 13(10), 1238-1254. <https://doi.org/10.1080/17512786.2019.1585198>
- XIE, X., GAI, X., ZHOU, Y. (2019). A meta-analysis of media literacy interventions for deviant behaviors. *Computers & Education*, 139, 146-156.
- ZIMMER, F.; SCHEIBE, K.; STOCK, M.; STOCK, W.G. (2019). "Fake news in social media: Bad algorithms or biased users?". *JISTaPb*, v. 7, n. 2, pp. 40-53. <https://doi.org/10.1633/JISTaP2019.7.2.4>

Analysis on disinformation and gender bias generated by artificial intelligence

Abstract

The convergence between artificial intelligence (AI), gender bias and disinformation has become a central theme in the debate on contemporary communication. This research comprehensively analyses how biased algorithms influence the perpetuation of inequalities, all while exploring AI'S potential to mitigate gender stereotypes and promote equity. It addresses the effects of disinformation generated by AI in the perception and representation of gender in different social contexts, highlighting its impact on power dynamics and building inclusive narratives. Additionally, the study examines the role of gender media literacy and education as key tools to counteract these issues. We analyse international initiatives and public policies intended to integrate a gender perspective into development and use of AI technologies, as well as feminist focuses in data science that question traditional structures and propose more inclusive alternatives. Finally, this research seeks to contribute to the interdisciplinary debate on gender and technology, highlighting the need for a collaborative focus to build fairer and more inclusive technological systems that benefit everyone, regardless of their gender.

Keywords

Artificial intelligence (AI), gender disinformation, gender biases, algorithmic discrimination, gender media literacy.

1. Introduction

Gender bias in AI is arising as one of the most urgent and debated issues in contemporary technological development. On one hand, recent research proves the marked underrepresentation of women in AI system design and development teams, and on the other, the consequences of this exclusion in perpetuating structural inequalities are of note (Leavy, 2018; Nadeem et al., 2020). These dynamics not only reflect historical inequalities, but also reveal how algorithmic systems amplify biases present in society.

One of the leading causes of this issue is the lack of diversity in data sets used to train automatic learning models, as well as the homogeneous composition of the teams responsible for creating them. Additionally, the implicit bias of the programmers, whether conscious or unconscious, translates to systems that reproduce gender stereotypes and limit the technology's transformative potential (Nadeem et al., 2020). This reality highlights the need to implement corrective measures, such as incorporating gender perspectives and fostering diversity, not only to enrich the functionality of the systems, but also to guarantee greater equity in their design and application (Leavy, 2018; Liu, 2024).

In this context, different strategies have been proposed to mitigate the impact of gender bias on AI. One of the most noteworthy strategies is the implementation of equity principles in developing algorithms, systematic reduction of the algorithmic bias, and a significant increase in women's participation in STEM disciplines (science, technology, engineering, and maths) (Nadeem et al., 2020; Liu, 2024). Moreover, ethical focuses such as distributive justice and the ethics of care are being integrated into the analysis and design of these systems, providing theoretical frameworks that promote the creation of more inclusive technologies that are sensitive to the needs of a diverse population (Wellner, 2020).

On the other hand, despite progress toward gender equality in different spheres of society, professional and social hierarchies are still mainly dominated by men, reflecting unequal power distribution (Latu et al., 2015). This highlights the urgency of tackling the structural inequalities that underlie technological and social dynamics, beginning by acknowledging that the concept of gender transcends physiological differences to address cultural and social constructs. As indicated by DCAF, Geneva Centre for Security Sector Governance, gender equality must be understood as a primary objective, which includes equalisation of rights, equal availability of resources and opportunities, and economic independence, regardless of gender.

Consequently, it is fundamental to highlight that in the fight against gender inequality, one of the clearest manifestations is discrimination against women. As established by the United Nations in its first article, this is defined as:

"distinction, exclusion or restriction made on the basis of sex which has the effect or purpose of impairing or nullifying the recognition, enjoyment or exercise by women, irrespective of their marital status, on a basis of equality of men and women, of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field" (UN: General Assembly, Convention on the Elimination of All Forms of Discrimination against Women, 1979).

Based on this definition, we can see how said discrimination manifests in different fields. For example, in recent years, the increased violence against women in digital settings is particularly alarming. Social media platforms have become spaces where misogynist discourse is not only generated, but also organised and amplified, becoming veritable campaigns of gender disinformation (Cabañes, 2020). This phenomenon proves that social structures that perpetuate inequalities and violence have found new mechanisms in digital technologies to invalidate or exclude minority social groups (Benjamin, 2019). Moreover, hate speech has intensified on these platforms, attacking individuals based on nationality, ethnicity, race, sexual orientation, and gender, and spreading with alarming speed (Lapa and Di Fátima, 2023).

However, these forms of discrimination are not only limited to the digital world. They also persist in other key fields, like education, where gender inequalities continue to affect millions of women. Particularly, women face significant barriers and prejudices in accessing primary education, especially in developing nations. In these regions, girls represent a considerable

proportion of those who rarely access basic schooling (UNESCO, 2018). This panorama highlights the need to implement early, sustained actions that not only address these inequalities, but that also foster equity and empowerment for women and girls in all spheres of society.

As an indispensable condition, gender equality and empowering women and girls require a guarantee for all of them of full access to their rights and fundamental freedoms. This principle is set forth in a resolution adopted by the Economic and Social Council dated 8 June 2015, which seeks to accelerate achievement of gender equality by implementing measures aligned with the Convention on the Elimination of All Forms of Discrimination against Women. Both initiatives mutually reinforce one another in their commitment to foster women's empowerment in all fields of social, economic, and cultural life (ECOSOC, 2015).

In this line, the Beijing Declaration emphasises the need to conduct specific actions that drive gender equality. They include fostering women's and gender studies, as well as applying the results obtained to key areas like the economy, science, and technology (Beijing Declaration). Notwithstanding, despite this progress, the path toward equity between men and women remains challenging. Persistent gender inequality, deeply rooted in fields like advertising, film, music videos, and television, takes on new dimensions in the digital field. This occurs when technological platforms integrate biased algorithmic decisions that not only perpetuate, but also amplify prejudices, creating a vicious cycle that is not always clear (Flexer et al., 2018; Gutiérrez, 2021; Hajian et al., 2016).

This dynamic has led to a relatively recent, but just as concerning, phenomenon: gender disinformation. Although studies on disinformation have garnered attention, especially after Donald Trump's presidential victory in 2016 (Freelon and Wells, 2020), and more recently in 2024, there is still much to research on how women and underrepresented groups become the target of attacks and on the long-term consequences of these practises.

Additionally, in technology, women continue to be significantly underrepresented. As indicated by Wajcman (2010), women are a minority in key sectors like information technology, electronics, and communications. While their participation in STEM education and careers (science, technology, engineering, and maths) has grown, the gender gap still considerably persists (Davila Dos Santos et al., 2022). According to UNESCO (2019), only 12% of AI researchers are women, and they represent barely 6% of software developers. Additionally, they are 13 times less likely to register an information and communication technologies patent (ICT) in comparison with their male colleagues. This degree of underrepresentation leads to the exclusion of women in technical design processes, which limits their influence on the configuration of these technologies (Wajcman, 2010, p. 145).

The impact of this exclusion can also be seen on the job market. A study conducted by the World Economic Forum (2018) indicates that 57% of jobs that could be replaced by technologies fall under women, disproportionately affecting jobs that belong to minorities who work in service sectors (World Economic Forum Boston Consulting Group [BCG], 2018; Zimmerman, 2018).

On the other hand, during the UNESCO'S Global Dialogue on Gender Equality and Artificial Intelligence, conducted in 2020, they identified a concerning lack of regulatory instruments to

effectively address gender equality as an independent issue. They concluded that current practises are insufficient to face this issue comprehensively (UNESCO, 2020, p. 12).

Given this scenario, the UNESCO'S "Women4Ethical AI" initiative called for global cooperation to develop an inclusive, gender-sensitive AI. At the World Forum on the Ethics of Artificial Intelligence, held in Slovenia in February 2024, the network issued a statement that highlighted the importance of aligning AI technologies with human rights principles. Proposed measures include the implementation of solid assessments, establishing accountability mechanisms, human supervision over algorithmic processes, ethical data management with a focus on rights, and integration of the gender perspective into all technological initiatives. These actions are an essential step toward development of more equal and responsible AI systems that promote true social and gender inclusion.

Although we are fully in the "Age of Algorithms" (Allen and Masters, 2020), the impact of these technologies on gender equality is still insufficiently explored (Alpaydin, 2021; Bolukbasi et al., 2016). This highlights the need for a more detailed analysis on the challenges and opportunities borne by algorithms, as well as possible discriminatory results that could arise from implementing them (Brière, 2022).

In the past, technology has been associated with the androcracy and is characterised as a male domain, where women were represented as inept and lacking in skill and interest, which perpetuates an exclusionary vision of their participation (Johnson, 2010; Onta, 2007). However, feminist interventions in technology have arisen as critical tools to question these narratives. These interventions are focused on examining how technologies reflect and reinforce gender constructs and the performance of traditional roles (Sutko, 2020, p. 568). From this perspective, gender relations are "materialised in technology," strengthening inequalities that are not only social, but also technical (Sutko, 2020, p. 568). Fryxell (2021) delves into this idea by arguing that digital worlds are "configured by gender technologies that ratify heteronormative roles and objectify women in the process" (p. 33). In this regard, AI with a gender perspective is a clear example of how gender divisions are "naturalised and (re)produce" on technological platforms (Sutko, 2020).

Feminism in data science is also gaining relevance as a field that challenges the supposed neutrality and objectivity of algorithmic systems. This focus is based on the idea that data systems are products of unequal social relationships, rather than impartial tools. Through the concept of intersectionality, D'Ignazio and Klein (2019) have shown how data feminism can be put into practise through principles that include questioning hierarchies, valuing emotion as a form of knowledge, and raising awareness of women's work, especially during the COVID-19 pandemic (D'Ignazio and Klein, 2020).

In this context, legislation and public policy play a key role in fighting disinformation and gender discrimination. These tools are essential to establish regulations on digital platforms and to fight online gender-based violence (Wilfore, 2021; Amaral, Simões and Poleac, 2022). However, as highlighted by Broussard (2016), analysing algorithms from a legal perspective is still an infrequent practise, especially in cases related to gender equality and discrimination (Buijsman and Jänicke, 2021).

Moreover, gender media literacy arises as an essential strategy to fight disinformation and discrimination. This practise fosters critical education on media and technologies, allowing both professionals and citizens to identify and address biases from a very early stage.

Consequently, the gender bias in AI not only poses a technical challenge, but also represents an ethical and social imperative. Addressing this issue will help to build fairer, more responsible systems which, ultimately, benefit a greater diversity of users and promote inclusive social progress (Liu, 2024).

In this research, we explore the intersection between disinformation, gender stereotypes, and the role of AI in perpetuating or mitigating these dynamics. Based on different hypotheses, we analyse how disinformation influences the perpetuation of gender stereotypes and how algorithmic biases affect the production and dissemination of biased information.

Moreover, we examine types of gender disinformation, from the distortion of information about men and women to the dissemination of fake news that reinforce stereotyped roles. We also consider how algorithmic biases impact the generation and consumption of online content, included the behaviour of virtual assistants.

Finally, we discuss existing regulatory policies and proposals to address these issues, highlighting the role of gender media literacy as a fundamental tool to empower people to identify and question biases in digital technologies. We present conclusions that highlight the practical and theoretical implications to fight gender disinformation and promote more inclusive communication in the digital world.

2. Research Hypothesis

Within the context of this research, we have formulated three fundamental research hypotheses to understand the dynamic between disinformation, AI, and gender biases:

1. There is a significant relationship between exposure to AI-generated disinformation and the perpetuation of gender stereotypes in society, which affects gender equality and fair treatment.
2. Gender biases in AI algorithms have a significant impact on women's opportunities for development and decision-making in different professional fields, which contributes to gender disparity in different areas.
3. There is a significant relationship between the degree of literacy and understanding of gender biases in AI and reduction of gender discrimination and disinformation in society, which promotes gender equality and equity.

To address these issues, we have conducted a systematised review of the bibliography based on key concepts like "*desinformación de género*," "*sesgos de género*," and "*Inteligencia Artificial*," as well as they synonyms and similar definitions in English (like "gender

disinformation," "disinformation," "gender bias," "gender equality," "femtech"), both in the field of journalism and in the media.

3. Analysis

3.1. Disinformation and perpetuation of gender stereotypes

Gender disinformation is a set of complex and deliberate practises designed to perpetuate structural inequalities and reinforce harmful stereotypes about women. These strategies not only seek to undermine equality, but also to strengthen misogynistic narratives that limit the participation and visibility of women in society. From the dissemination of non-consensual sexual content to manipulation of private information, these actions have profound consequences on the safety and well-being of women in the digital world.

Specifically, gender disinformation can be defined as the circulation of fake or deceitful information that attacks women based on their gender identity, with a particular focus on female political leaders, journalists, and public figures. This kind of attack is based on narratives deeply-rooted in stereotyped gender roles, whose object is to delegitimise their achievements and limit their influence (Jankowicz, 2017; Barker and Jurasz, 2019; Di Meco, 2019; Stabile et al., 2019). The techniques used to spread this disinformation include misogynistic comments that reinforce traditional stereotypes, sexualisation through manipulated images, online harassment, and cyber-attacks. These practises generate a negative perception of women in society, eroding their credibility in leadership roles, discouraging their participation in public debates and promoting their silencing (Di Meco, 2020).

Some of the leading victims of these campaigns are high-profile women, such as celebrities, activists, political leaders, journalists, and researchers. These figures face constant and recurring harassment, which seeks not only to discredit their work, but also to intimidate them into leaving their public positions (Marwick and Caplan, 2018; Stabile et al., 2019; Chen et al., 2020; Murphy and Flynn, 2022; Di Meco, 2023). This kind of disinformation perpetuates the idea that women are unreliable or excessively emotional to hold roles of power, especially in the political sphere, contributing to the exclusion of their voices in decision-making spaces. According to Di Meco and Wilfore (2021), this disinformation strategy's effect is to discourage women from seeking out leadership positions, perpetuating their underrepresentation in high-visibility fields. As mentioned by Lucina Di Meco (2020, p. 4), gender disinformation implies "spreading deceitful or inaccurate information and images against female political leaders, journalists, and public figures, based on misogyny and stereotyped gender roles." This phenomenon is also related to "dark participation," where women are represented as participants in hidden or manipulative agendas to justify their exclusion (Quandt, 2018).

Specifically, and in the political sphere, for example, disinformation narratives are systematically used to weaken the authority of women leaders. Several studies have proven that these strategies contribute to the underrepresentation of women in politics by associating them with stereotypes that reduce their leadership capabilities (Hedlund et al., 1979; Dahlerup, 2006; Elder, 2008). Female leaders tend to be the target of superficial criticisms that reduce their initiatives to aesthetic or irrelevant issues, while belittling their contributions to complex

topics. For example, Garikipati and Kambhampati (2020) analysed responses to the COVID-19 pandemic in countries led by women, like New Zealand, Taiwan, and Germany, concluding that their decisions reflected greater risk aversion in terms of human loss, which contrasts with the priority granted to economic considerations in other contexts. Although this participatory and empathetic focus has been recognised as a strength, it has also been interpreted by some as a sign of weakness of excessive sensitiveness (Eagly and Johnson, 1990).

Moreover, the trend to infantilise women reinforces the image that they are incapable of making complex decisions or holding leadership roles (Carlson, 2010; Huot, 2013). This narrative is shown in disinformation strategies, where existing narratives are overlapped and recycled to provoke emotional responses in audiences. This phenomenon is driven by the complex nature of disinformation systems and the tendency of those who believe in a manipulated narrative to accept other similar narratives, thereby amplifying their impact (Goertzel, 1994).

3.2. Types of gender disinformation

Gender disinformation is a broad phenomenon that uses different methods and techniques to perpetuate gender stereotypes and inequalities. One of the most well-known forms of this disinformation is "revenge porn," which consists of distributing sexual content without the consent of the party involved. This practise implies the use of doctored images or private information in order to extort or harass the victims, perpetuating a cycle of abuse and control, and also causing a devastating impact on their physical and emotional well-being (Gámez-Guadix et al., 2022).

This violates not only the privacy and autonomy of the victims, but also their emotional well-being and mental health, exposing them to serious psychological and social consequences (Hearn and Hall, 2019).

Another manifestation of this phenomenon is "creepshots," "upskirting," or "digital voyeurism," which means capturing and distributing images or videos of individuals in intimate or vulnerable situations without their consent. These practises, motivated by voyeurism or the intent to degrade the individuals in the images, reinforce the objectification of women and perpetuate gender violence, undermining the victims' trust and safety in the digital world (Lewis and Anitha, 2023).

"Doxing," which means publishing or doctoring personal information without consent, is also a recurring tactic in gender disinformation (Chen et al., 2019). By revealing sensitive information, these practise seeks to intimidate and harass the victims, placing their physical and emotional integrity at risk. This proves the urgent need to address this issue from both a legal and an ethical perspective.

Impersonation is another tool used to spread gender disinformation, especially within contexts of threats and defamation online. By stealing the individual's identity, the perpetrators seek to harm their reputation and credibility, which can lead to devastating repercussions, both in their personal and in their professional life (Cross and Layt, 2022).

On the other hand, "mobbing," which consists of mobilising a group to harass or intimidate an individual online, is also a method of gender disinformation worthy of attention. This kind of harassment, frequently deployed on social media, can severely affect the mental and emotional health of victims, limiting their ability to actively participate in the public sphere (Topkaya, 2011).

Sexist hate speech is another key mechanism in perpetuating harmful stereotypes and inciting violence against women. This speech, which often presents women as sexual objects or targets of violence, contributes to a culture of misogyny and discrimination that affects all levels of society (Wojatzki, 2018).

However, not all gender disinformation tactics are clearly identified, since many of them are presented through strategies that are less visible, but just as harmful. For example, one of these strategies is deliberate confusion between domestic violence and gender-based violence, which hinders efforts to address and prevent gender-based violence effectively.

Another recurring narrative is the false affirmation that there is a high percentage of gender-based harassment reports that are false. Although these reports only represent a minimum fraction of the total, these narratives seek to delegitimise the experiences of the victims and trivialise gender-based violence. At the same time, by equating acquittals with false reports, the perpetrators of these narratives contribute to minimise the real impact of gender-based violence.

Moreover, the affirmation is also spread that men are also victims of domestic violence, which distorts gender realities and perpetuates a narrative of equity that is not supported by empirical evidence. Similarly, the erroneous idea that most feminicides are committed by foreign men foments ethnic stigmatisation and feeds into racist and xenophobic attitudes in society (Smith and Bamberger, 2021).

Together, these gender disinformation strategies help to maintain unequal power structures and hinder the construction of a fairer society. Their impact is not only limited to direct victims, but also negatively influences social progress and the perception of gender equality as an attainable objective.

3.3. Impact of gender biases on AI algorithms

The proliferation of applications, virtual assistants, and robots that use AI has brought with it the growing concern of gender bias, an issue that has taken on relevance both within the technological and contemporary social field. This phenomenon has intensified due to the mass implementation of these tools at different organisations, which has shown an increase in incidents related to gender bias. By mirroring and amplifying pre-existing social inequalities, AI machines highlight the importance of urgently addressing this issue (UNESCO, 2020; Stanovsky et al., 2019).

Data play a central role in the creation of AI algorithms, since data's quality and diversity are essential to guarantee optimum performance and fairness in results (Downey, 2021). However, the inherent prejudices, whether conscious or unconscious, of those who collect

and prepare these data can be integrated into the algorithms, causing discriminatory effects (Bui, 2021). To mitigate these issues, it is crucial to adopt rigorous ethical procedures and promote greater awareness of the social implications of the technology, applying training techniques designed to reduce bias (Kelley and Ovchinnikov, 2020).

The AI's quality depends directly on the input data: if they contain biases or are incomplete, the technology can amplify pre-existing inequalities. For this reason, it is essential to guarantee transparency in access to data, encourage auditing them, and correct possible biases during algorithmic development (Daley, 2021). This focus should also be used to address gender disparities in designing and implementing AI, a critical issue given the historical lack of diversity in this regard (Coeckelbergh, 2019; Equinet, 2021). Widely documented cases of algorithmic bias (Crawford, 2013; Eisenstat, 2019; Hao, 2019; Knight, 2016; Rodríguez Martínez and Gaubert, 2020; Wang, 2018) highlight the severity of the problem, while reports such as the ones by Data2x (2021) specifically examine how data affect the lives of women (Vaitla et al., 2017).

Algorithms, intended as tools to transform data into results, require huge volumes of information to learn, which makes them a direct reflection of the input data (Guszcza, 2018). Subjective decisions made while collecting and preparing these data can introduce significant biases (F. R. A. Focus, 2018), which reinforces the need to design systems that consider the social and contextual realities of the environment where they are implemented (Guszcza, 2018). Moreover, the lack of diversity in the AI community has contributed to the creation of algorithms which, from the very beginning, reinforce pre-existing inequalities (Equinet, 2021).

A noteworthy problem is the perpetuation of discriminatory biases and stereotypes through search algorithms, as observed in the images that mainly associate "CEO" with men (Jean, 2021). This kind of result influences social perceptions and can feed discriminatory decisions. Similarly, neuronal networks trained with biased data tend to magnify existing inequalities, like in the case of an algorithm designed to identify human activities in images, which developed gender biases due to the nature of the data used (Knight, 2019; Kleinberg et al., 2020).

On the other hand, Benjamin (2020) warns how AI systems offer "the appearance of objectivity without public responsibility" (p. 53), which shows the importance of paying attention to the collection, storage, and processing of data, as well as ethical and technical standards applied to algorithm design (Gilli et al., 2019). Designing human-focused technologies can help to address these issues, ensuring that they reflect the social contexts within which they are implemented and do not reinforce existing inequalities.

In this context, OpenAI has announced that its Sora application is equipped with protection mechanisms designed to prevent the generation of violent or sexual content or that imply use of someone's identity without their consent. Based on technologies like DALL-E, Sora promises to reject instructions that violate the policies of use. However, previous research has proven that these safety measures can be violated with specific instructions, which raises questions about the effectiveness of these protections and the possible exploitation of these tools for undue purposes (OpenAI, 2021). Moreover, there is a risk that these applications develop inherent biases or are used to commit fraud, highlighting the importance of establishing ethical principles and robust safety measures when developing AI technologies (Mehrabi et al., 2019).

Biased algorithmic decision-making has tangible repercussions on different fields of everyday life. One representative case is a European bank which, in granting car loans, demonstrated marked gender inequality. Despite having lower default rates, women faced greater difficulty in obtaining loans (Andreeva and Matuszyk, 2018). This example illustrates how algorithmic decision-making can intensify pre-existing social inequalities, highlighting the urgency of addressing these problems from an ethical and regulatory perspective (Hajian et al., 2016; Langston, 2015; Wachter-Boettcher, 2017; Zhao et al., 2017).

In 2015, Google apologized after its photo classification algorithm confused images of black humans with gorillas, an error that proved the deep limitations and biases in artificial intelligence systems. Two years later, in Arkansas, it was revealed that an algorithm used to manage access to medical benefits under the state's disability programme did not consider critical conditions like cerebral paralysis or diabetes, which led to significant concerns regarding justice and fairness in designing these systems.

Examples like these highlight the persistence of discrimination in using AI, a widely documented concern in academic studies. For example, error rates in gender classification are significantly higher for women and people of colour, which shows the presence of racial and gender bias in algorithmic systems. An emblematic case is the Amazon algorithm, which learned to devalue applications that included the word "women" and, at the same time, favoured CVs that contained words associated with male engineers, which indicated an implicit preference for men (Larry, 2018; Meyer, 2021).

Concerning examples of algorithmic discrimination have also been identified in the field of advertising. Job adverts, for example, tend to promote higher salaries for men than for women, reflecting gender bias in audience segmentation and in establishing advertising goals (Gibbs, 2015; Datta et al., 2015). Although advertising is just one of the many fields wherein these biases are clear, other sectors, such as facial recognition and hiring processes, also cause increasing concern. The algorithms used in these contexts can incorporate implicit prejudices that encourage certain skills or knowledge to the detriment of others, which can lead to discriminatory decisions (Greenfield and Griffin, 2018). To mitigate these problems, it is crucial to develop human-focused algorithms and foster ethical practises in designing AI technologies (Lambrecht and Tucker, 2016).

Bias also manifests in the news media, where cultural stereotypes are perpetuated that limit the roles of women to support figures, as assistants, caretakers, or teachers (Brody and Hall, 2008). In Italy, for example, it was observed that 58% of verified disinformative audio messages between March and May were delivered with female voices, which reinforces pre-existing gender roles and limits equitable representations.

Gender algorithmic bias is certainly a complex, persistent problem that affects different fields of life, from medical care and employment to advertising and financial decisions. The lack of transparency in algorithmic systems and the lack of diversity in the technological industry contribute to the perpetuation of these biases. For this reason, it is fundamental to implement policies and regulations that promote gender equity in developing and applying AI algorithms, guaranteeing that these technologies work for the benefits of a more inclusive, fairer society.

3.4. Virtual Assistants, AI, and Regulatory Measures

The relationship between women's work and technology has taken on growing interest in gender and technology studies. Thus, a socialist feminist focus, as pointed out by Wajcman (2010) provides a valuable framework to analyse intersections between women's work and AI. This focus highlights how the "domestication" of AI is built around an association with femininity, seen as docile, receptive, and oriented toward caretaking (Sutko, 2020).

An illustrative example of this idea is found in virtual assistants like Siri, Cortana, and Alexa, who symbolise what is called "feminisation of AI." These devices normalise dividing work by gender by associating femininity with symbolic and communicative tasks (Sutko, 2020). These technologies are designed to conduct roles that were traditionally related to female gender roles, such as calendar management, note taking, and task organisation (Sutko, 2020).

Moreover, it should be noted that most of these virtual assistants have female voices, which reinforces gender stereotypes and perpetuates the feminisation of technology. Although Google Assistant is an exception, since it does not have a name associated with a specific gender, devices like Siri, Alexa, and Cortana have feminine voices, highlighting the tendency to relate femininity with communication and customer services roles (Costa & Ribas, 2019; UNESCO, 2019).

This feminisation of AI also reinforces traditional gender norms through the performance of "emotional labour." This kind of work, which produces emotional responses and interpersonal relationships, is a central characteristic of post-Fordist economies dominated by the service sector (Altomonte, 2015). Said emotional labour, often compared with caretaking and domestic work, contributes to perpetuating traditional gender roles (Hardt, 1999).

On the other hand, gender preferences are also clear in the use of male voices in contexts perceived as authoritarian or professional. For example, IBM's Watson, used in medical settings, uses a male voice, reflecting the perception that these voices are more appropriate for professional and instructional contexts (Costa & Ribas, 2019).

Objectifying women also takes on new shapes through AI, when materialised through gendered robots. A paradigmatic case is Sophia, a humanoid robot developed by Hanson Robotics and designed to be "exceptionally attractive" and evoke a sensation of "mechanic-eroticism" (Zimmerman, 2018). This trend also manifests in selecting male and female robots for different occupational roles, perpetuating gender stereotypes and inequalities in the professional field (Nomura, 2020; Eyssel & Hegel, 2012).

To move toward more inclusive, fairer technology, it is essential to comprehensively address gender biases in the development and use of AI. In this regard, the European Union's proposal regarding the Artificial Intelligence Act (AIA) introduces significant measures to fight gender biases and promote equality in development and application of AI.

The AIA establishes regulatory restrictions and actions that directly address issues related to gender equality. This shows recognition on the European Union's part of the need to mitigate gender biases in AI and complements the existing legal framework with specific provisions on gender equality and non-discrimination (European Commission, 2022).

One of the AIA's key elements is its focus on the algorithm design phase. The regulation establishes that data sets used for training, validating, and testing algorithms must be representative, complete, and error-free, in order to prevent the incorporation of gender bias from the beginning of technological development. Additionally, it includes data management and governance practises to identify and correct possible biases during these stages (European Commission, 2022).

This proactive focus marks a significant change in development and use of AI, fostering greater scrutiny of gender biases and promoting more inclusive practises within the technological industry. Moreover, the AIA establishes transparency and monitoring obligations for after commercialising AI systems. This includes the creation of records during the training phase and later implementation, guaranteeing system traceability and facilitating the investigation of possible malfunctions or inappropriate behaviours (European Commission, 2022).

3.5. Relationship Between Gender Media Literacy, Disinformation, and Gender Discrimination

AI has taken its position as a key tool in the digital transformation, but it also faces criticism for perpetuating gender biases present in the data and algorithms underlying it. The Organisation of Consumers and Users (OCU) has identified examples of these biases in tools like ChatGPT and Perplexity, highlighting the urgent need to design algorithms that do not reinforce social inequalities (Organisation of Consumers and Users, 2024). Moreover, recent research reveals that AI systems do not only reflect social prejudices, but can also amplify them, generated chauvinism- and racism-laden content that reinforces negative stereotypes (Cabrera, 2024).

Gender inequality, now recognised as a factor that feeds violence in multiple social spheres, finds a channel to spread through digital technologies. Studies like the ones by Liberia, Zurbano, and Barredo (2013), Estébanez and Vázquez (2013), and Muñiz and Cuesta (2015) show how these phenomena are aggravated in the digital age. Additionally, based on Díaz-Aguado's (2013) methodology, Rosser et al. (2015) analyses the relationship of gender and sexist content on social media, highlighting the magnitude of the problem. Along with platforms like WhatsApp, these platforms not only disseminate traditional gender roles, but also legitimise sexist microaggressions, many times rendered invisible by their implicit social acceptance. For this reason, strengthening the critical ability of users is indispensable to detect and fight these messages (Aguaded Gómez and Díaz Gómez, 2008).

Moreover, gender disinformation is an urgent challenge, given that it disproportionately affects women and has serious consequences for democracy and political participation. According to Atuase (2018), misogyny and other biases present in disinformation perpetuate stereotypes and feed into discriminatory dynamics. Therefore, fostering critical thought and actively questioning sources and messages is an essential strategy to counteract this phenomenon (Sessa, 2020).

In this context, gender media literacy is an essential tool to promote a fairer society. In addition to fighting disinformation and discrimination, it fosters inclusive education that incentivises a

critical look at media contents, laying the foundation for a new generation of news professionals who are committed to challenging gender stereotypes. This focus, supported by UNESCO, highlights the importance of generating content free from prejudice and inequality (García Matilla, 2015; López Safi, 2015; Buitrago, Navarro and García Matilla, 2015).

Only by means of a comprehensive focus that addresses education, technological ethics, and media literacy will it be possible to build a society that rejects inequality and promotes gender equity.

4. Conclusions:

This research highlights how gender disinformation and biases in artificial intelligence perpetuate structural inequalities and stereotypes, affecting both women's representation and their access to professional opportunities and fundamental rights. These issues, deeply rooted in current social and technological dynamics, highlight the importance of addressing them comprehensively, joining educational, legislative, and technological efforts.

In relation to the hypotheses provided, our findings confirm the following:

1. Relationship between algorithmic disinformation and perpetuation of gender stereotypes: Exposure to disinformation generated by AI reinforces stereotyped narratives and roles, negatively affecting gender equality. This translates to the consolidation of discriminatory narratives that impact both public perception and women's ability to hold leadership roles or participate in the public sphere.
2. Impact of gender biases in algorithms on professional opportunities: The biases identified in algorithmic systems affect hiring processes, resource allocation, and advertising segmentation, which significantly limits women's possibilities for professional development. This phenomenon perpetuates gender disparities in key sectors, including STEM disciplines and emerging technologies.
3. Importance of gender media literacy to reduce discrimination: Critical education on gender bias in AI and media literacy empower individuals to identify and question prejudices in technological systems. This focus helps to reduce discrimination and disinformation, fostering a more inclusive society.

Based on these findings, we can reflect on future implications and the path toward more inclusive development. One of the most urgent aspects is the need to guarantee that artificial intelligence systems reflect values of equity and social justice instead of reproducing inequalities from the past. To achieve this, it is fundamental to incorporate gender and intersectional perspectives from the early stages of technology design and development. This integration will not only enrich the systems created but will also contribute to society trusting in these tools.

Additionally, public policies must evolve to adapt to the complexity of current challenges. It is essential that governments adopt regulatory frameworks that not only regulate AI systems, but also foster the active participation of underrepresented groups in the technological industry. This way, the homogeneity of development teams can be reduced, which is one of the leading causes of algorithmic biases.

The role of education is also a central cornerstone for this transformation. Training in digital skills and media literacy with a gender perspective should be promoted from an early age and spread to professional contexts. Only through informed citizenry will it be possible to counteract the harmful effects of disinformation and foster critical and ethical use of technology.

On the other hand, inter-sectorial collaboration is an essential strategy to address these issues. Academic institutions, technological companies, non-governmental organisations, and international bodies must work together to develop innovative solutions that mitigate gender biases and fight disinformation. This cooperation can also facilitate the creation of global standards that guarantee transparency, fairness, and accountability in the use of artificial intelligence.

Finally, it is crucial to reflect on the ethical role of technology in society. Artificial intelligence is not an autonomous entity; it is a reflection of the human decisions that shape it. Therefore, the true potential of these tools lies in their ability to be responsibly designed, prioritising diversity and respect for fundamental rights. Eventually, overcoming the challenges posed by gender disinformation and biases in AI will not only improve technological systems, but will also contribute to building a fairer and more inclusive society, where every individual, regardless of their gender, can participate and thrive under equal conditions.

References:

Aguaded Gómez, José Ignacio y Díaz Gómez, Rocío (2008): "La formación de telespectadores críticos en educación secundaria", en *Revista Latina de Comunicación Social* 63, 121-139. doi: <http://goo.gl/r8rZin>

Allen, R., Masters, D.: Artificial intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making. *ERA Forum* 20, 585–598 (2020)

Alpaydin, E.: *Machine Learning*. Mit Press, Boston (2021)

Altomonte, G. (2015). Affective labor in the post-fordist transformation. Public Seminar.

Amaral, Inês, Rita Basílio Simões, & Gabriela Poleac. 2022. "Technology gap and other tensions in social support and legal procedures: stakeholders' perceptions of online violence against women during the Covid-19 pandemic." *Profesional de la información* 31(4): e310413. DOI: <https://doi.org/10.3145/epi.2022.jul.13>

Andreeva, G., y Matuszyk, A. (2018). Gender discrimination in algorithmic decision-making. 2nd International Conference on Advanced Research Methods and Analytics (CARMA2018). <https://riunet.upv.es/bitstream/handle/10251/111932/8312-23285-1-PB.pdf?sequence=1&isAllowed=y>

Atuase, Diana. "Gender Equality and Women Empowerment in Ghana, the Role of Academic Libraries." *Journal of Applied Information Science* 6, no. 2 (2018): 14–20

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity.

Benjamin, R. (2020). Race after technology: Abolitionist tools for the new JIM code. *Social Forces*, 98(4), 1–3. <https://doi.org/10.1093/sf/soz162>

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., Kalai, A.: Man is to computer programmer as woman is to homemaker? Debiasing word embeddings (2016). arXiv:1607.06520

Bonnie Stabile, Aubrey Grant, Hemant Purohit & Kelsey Harris (2019) Sex, Lies, and Stereotypes: Gendered Implications of Fake News for Women in Politics, *Public Integrity*, 21:5, 491-502, DOI: 10.1080/10999922.2019.1626695

Brière, C., Dony, M.: *Droit de l'Union européenne*. Editions de l'Université de Bruxelles, Bruxelles (2022)

Brody, L. R., & Hall, J. A. (2008). Gender and emotion in context. *Handbook of emotions*, 3, 395-408.

Broussard, M.: *Artificial Unintelligence: How Computers Misunderstand the World*. Mit Press, Boston (2018)

Bui, Cathrine Kieu Trang. "Exploring Bias against Women in Artificial Intelligence: Practitioners' Views on Systems of Discrimination." DUO, September 26, (2021). <https://www.duo.uio.no/handle/10852/88551>.

Buijsman, S., Jänicke, B.: *Ada und die Algorithmen: wahre Geschichten aus der Welt der künstlichen Intelligenz*. C.H. Beck, München (2021)

Buitrago, A., Navarro, E. y García Matilla, A. (2015). *La educación mediática y los profesionales de la comunicación*. Barcelona: Gedisa.

Cabañes, Jason Vincent A. 2020. "Digital disinformation and the imaginative dimension of communication." *Journalism and Mass Communication Quarterly* 97(2): 435-452. DOI: <https://doi.org/10.1177/1077699020913799>

Cabrera, C. (2024, 9 de diciembre). La IA genera audios plagados de machismo, racismo e infracciones de derechos de autor. *El País*. Recuperado de <https://elpais.com/tecnologia/2024-12-09/la-ia-genera-audios-plagados-de-machismo-racismo-e-infracciones-de-derechos-de-autor.html>

Carlson, C. (2010). Desensitization of infantilization. *Journal of Undergraduate Research*, 13, 1-23.

Chen, Gina Masullo, et al. 2020. "'You really have to have a thick skin': A cross-cultural perspective on how online harassment influences female journalists." *Journalism* 21(7): 877-895. DOI: <https://doi.org/10.1177/1464884918768500>

Chen, M., Cheung, A. S. Y., & Chan, K. L. (2019). Doxing: What adolescents look for and their intentions. *International journal of environmental research and public health*, 16(2), 218.

Coeckelbergh, Mark. "Technology Games/Gender Games. from Wittgenstein's Toolbox and Language Games to Gendered Robots and Biased Artificial Intelligence." *Techno:Phil – Aktuelle Herausforderungen der Technikphilosophie*, (2019), 27–38. https://doi.org/10.1007/978-3-476-04967-4_2.

Consiglio di Stato, sentenza n. 7891 del 4-25 novembre 2021. Available at: <https://www.eius.it/giurisprudenza/2021/655>

Costa, P., & Ribas, L. (2019). AI becomes her: Discussing gender and artificial intelligence. *Technoetic Arts*, 17(1), 171–193. https://doi.org/10.1386/tear_00014_1

Crawford, K. (2013). The Hidden Biases in Big Data. *Harvard Business Review*. <https://hbr.org/2013/04/the-hidden-biases-in-big-data>

Criado Perez, C.: We need to close the gender data gap by including women in our algorithms. *Time Magazine* 16.1.2020, (2020). Available at: <https://time.com/collection-post/5764698/genderdata-gap/>

Cross, C., & Layt, R. (2022). "I suspect that the pictures are stolen": Romance fraud, identity crime, and responding to suspicions of inauthentic identities. *Social Science Computer Review*, 40(4), 955-973.

Dahlerup, D. (Ed.). (2005). *Women, Quotas and Politics* (1st ed.). Routledge. <https://doi.org/10.4324/9780203099544>

Daley, Lauren Pasquarella. "Ai and Gender Bias (Trend Brief)." *Catalyst*, December 15, 2021. <https://www.catalyst.org/research/trend-brief-gender-bias-in-ai/>

DALL·E 3. (n.d.). <https://openai.com/dall-e-3>

Datta, Amit, Michael Carl Tschantz, and Anupam Datta. 2015. "Automated Experiments on Ad Privacy Settings." *Proceedings on Privacy Enhancing Technologies* 2015 (1): 92–112. <https://doi.org/10.1515/popets-20150007>.

Davila Dos Santos, E., Albahari, A., Díaz, S., & De Freitas, E. C. (2022). Science and Technology as Feminine': Raising awareness about and reducing the gender gap in STEM careers. *Journal of Gender Studies*, 31(4), 505–518. <https://doi.org/10.1080/09589236.2021.1922272>

De La Baume, M.: Germany to back EU's women quota plan after a decade (2022). Available at: <https://www.politico.eu/article/germany-will-adopt-women-on-board-directive-eu-proposal-after10-years-of-deadlock/>

Denny, J. (2020). What is an algorithm? How computers know what to do with data. *The Conversation*.

Díaz-Aguado. (2013). *La juventud universitaria ante la desigualdad y la violencia de género*. Madrid: Ministerio de Sanidad, Servicios Sociales e Igualdad.

Di Meco, Lucina. 2020. "Online Threats to Women's Political Participation and The Need for a Multi-Stakeholder, Cohesive Approach to Address Them." In Sixty-Fifth Session of the Commission on the Status of Women, 5-8 October. New York: UN Women Expert Group Meeting. Available on https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/CSW/65/EGM/DiMeco_OnlineThreats_EP8_EGMCSW65.pdf

Di Meco, Lucina. 2023. "Monetizing misogyny: Gendered disinformation and the undermining of women's rights and democracy globally." #ShePersisted. Available on https://she-persisted.org/wp-content/uploads/2023/02/ShePersisted_MonetizingMisogyny.pdf

Di Meco, Lucina, & Kristina Wilfore. 2021. "Gendered Disinformation Is a National Security Problem." Brookings TechStream. Available on <https://www.brookings.edu/techstream/gendered-disinformation-is-a-national-security-problem/>

D'Ignazio, C., y Klein, L. F. (2019). Data feminism. MIT Press.

D'Ignazio, C., y Klein, L. F. (2020). Seven intersectional feminist principles for equitable and actionable COVID-19 data. *Big Data & Society*, 7(2), 1-6. <https://doi.org/10.1177/2053951720942544>

Downey, L. (2021). Algorithms. Investopedia.

Eagly, Alice H. and Johnson, Blair T., "Gender and Leadership Style: A Meta-Analysis" (1990). CHIP Documents. 11. https://opencommons.uconn.edu/chip_docs/11

Eisenstat, Y. (2019). The Real Reason Tech Struggles with Algorithmic Bias. *Wired*. <https://www.wired.com/story/the-real-reason-tech-struggles-with-algorithmic-bias/>

Equinet, Contribution to the public consultation of the AIA (2021). Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legalrequirements/details/F2665651_en

Estébanez, I. y Vázquez, N. (2013). La desigualdad de género y el sexismo en las redes sociales. Una aproximación cualitativa al uso que hacen de las redes sociales las y los jóvenes de la CAPV. *¿Observatorio Vasco de la Juventud?*, Vitoria-Gastéiz: Servicio Central de Publicaciones del Gobierno Vasco.

European Commission, Better Regulation (2022). Available at: https://ec.europa.eu/info/law/lawmaking-process/planning-and-proposing-law/better-regulation-why-and-how_de#internationalzusammenarbeit-in-regulierungsfragen

European Law Institute, Artificial Intelligence (AI) and Public Administration – Developing Impact Assessments and Public Participation for Digital Democracy (2022). Available at: <https://www.europeanlawinstitute.eu/projects-publications/completed-projects-old/ai-and-public-administration/>

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology*, 42(9), 2213–2230. <https://doi.org/10.1111/j.1559-1816.2012.00937.x>

Flexer, A., Doerfler, M., Schluter, J., y Grill, T. (2018). Technical Algorithmic Bias. In A Music Recommender. 19th International Society for Music Information Retrieval Conference. <http://www.ofai.at/cgi-bin/get-tr?download=1&paper=oefai-tr-2018-03.pdf>

F. R. A. Focus. (2018). BigData: Discrimination in data-supported decision making. European Union Agency Fundamental Rights, Vienna, Austria, Rep. TK-02-18-634-EN-N.

Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political communication*, 37(2), 145-156.

Fryxell, A. R. (2021). Artificial Eye: The Modernist Origins of AI's Gender Problem. *Discourse*, 43(1), 31–64. <https://doi.org/10.13110/discourse.43.1.0031>

Gámez-Guadix, M., Mateos-Pérez, E., Wachs, S., Wright, M., Martínez, J., & Íncera, D. (2022). Assessing image-based sexual abuse: Measurement, prevalence, and temporal stability of sextortion and nonconsensual sexting ("revenge porn") among adolescents. *Journal of Adolescence*, 94(5), 789-799.

García Matilla, A. (2015). Una televisión para la educación. La utopía posible. Barcelona: Gedisa.

Gendered Misinformation & Online Violence Against Women in Politics: Capturing legal responsibility? | Co-inform. (n.d.). <https://coinform.eu/gendered-misinformation-online-violence-against-women-in-politics-capturing-legal-responsibility/>

Gibbs, Samuel. "Women Less Likely to Be Shown Ads for High-Paid Jobs on Google, Study Shows." *The Guardian*. Guardian News and Media, July 8, 2015. <https://www.theguardian.com/technology/2015/jul/08/women-less-likelyads-high-paid-jobs-google-study>.

Glick, P., & Fiske, S. T. (2001). Ambivalent sexism. In *Advances in experimental social psychology* (Vol. 33, pp. 115-188). Academic Press.

Goertzel, T. (1994). Belief in Conspiracy Theories. *Political Psychology*, 15(4), 731–742. <https://doi.org/10.2307/3791630>

Greenfield, G., McPherson, S., Mills, K., & McMullin, M. F. (2018). The ruxolitinib effect: understanding how molecular pathogenesis and epigenetic dysregulation impact therapeutic efficacy in myeloproliferative neoplasms. *Journal of translational medicine*, 16(1), 360.

Guszcza, J. (2018). Smarter together: Why artificial intelligence needs human-centered design. Deloitte. <https://www2.deloitte.com/us/en/insights/deloitte-review/issue-22/artificial-intelligence-human-centric-design.html>

Gutiérrez, M. (2021). Algorithmic Gender Bias and Audiovisual Data: A Research Agenda. *International Journal of Communication*, 15, 439-461. <https://ijoc.org/index.php/ijoc/article/viewFile/14906/3333>

- Hajian, S., Bonchi, F., y Castillo, C. (2016). Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. The 22nd ACM SIGKDD International Conference, Rec. <https://doi.org/10.1145/2939672.2945386>
- Hao, K. (2019). This is how AI bias really happens—And why it's so hard to fix. MIT Technology Review. <https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/>
- Haraldsson, A., & Wängnerud, L. (2019). The effect of media sexism on women's political ambition: evidence from a worldwide study. *Feminist media studies*, 19(4), 525-541.
- Hardesty, Larry. "Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems." MIT News | Massachusetts Institute of Technology, February 11, 2018. <https://news.mit.edu/2018/study-finds-gender-skin-typebias-artificial-intelligence-systems-0212>.
- Hardt, M. (1999). Affective labor. *Boundary*, 26(2), 89–100. <http://www.jstor.org/stable/303793>
- Hearn, J., & Hall, M. (2019). 'This is my cheating ex': Gender and sexuality in revenge porn. *Sexualities*, 22(5-6), 860-882.
- Hedlund, R. D., Freeman, P. K., Hamm, K. E., & Stein, R. A. (1979). The electability of women Candidates: The effects of sex role stereotypes. *The Journal of Politics*, 41(2), 513–524. <https://doi.org/10.2307/2129776>
- Huot, Chelsae R., "Language as a social reality: The effects of the infantilization of women" (2013). Dissertations and Theses @ UNI. 72.
- Jankowicz, N. (2022, March 31). How disinformation became a new threat to women. Coda Story. <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/>
- Jean, A.: *Les algorithmes font-ils la loi?* Observatoire (Editions de l'), Paris (2021)
- Johnson, D. (2010). *Sorting out the Question of Feminist Technology*. University of Illinois Press.
- Lamola, M. J. (2021). An ontic–ontological theory for ethics of designing social robots: a case of Black African women and humanoids. *Ethics and Information Technology*, 23(2), 119–126. <https://doi.org/10.1007/s10676-020-09529-z>
- Kelley, S., & Ovchinnikov, A. (2020). Anti-Discrimination Laws, AI, and Gender Bias in Non-Mortgage Fintech Lending. *SSRN Electronic Journal*.
- Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C.R.: Algorithms as discrimination detectors. *Proc. Natl. Acad. Sci.* 117(48), 30096–30100 (2020)
- Knight, W. (2016). How to Fix Silicon Valley's Sexist Algorithms: Computers are inheriting gender bias implanted in language data sets—And not everyone thinks we should correct it. MIT Technology Review. <https://www.technologyreview.com/s/602950/how-to-fix-silicon-valleys-sexist-algorithms/>

Knight, W.: The Apple Card Didn't 'See' Gender - and That's the Problem, the way its algorithm determines credit lines makes the risk of bias more acute, *Wired* 19.11.2019 (2019). Available at: <https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/>

Lambrecht, A., & Tucker, C. E. (2018). Field experiments. In *Handbook of Marketing Analytics* (pp. 32-51). Edward Elgar Publishing.

Langston, J. (2015). Who's a CEO? Google image results can shift gender biases. University of Washington. <https://www.washington.edu/news/2015/04/09/whos-a-ceo-google-image-results-can-shift-gender-biases/>

Lapa, Tiago, & Branco Di Fátima. 2023. "Hate Speech among security forces in Portugal." In *Hate Speech on Social Media*, edited by Branco Di Fátima, 277-293. Covilhã: LabCom Books.

Latu, Ioana, and Marianne Schmid Mast. "The Effects of Stereotypes of Women's Performance in Male-Dominated Hierarchies: Stereotype Threat Activation and Reduction through Role Models." *Gender and Social Hierarchies*, (2015), 87–99. <https://doi.org/10.4324/9781315675879-15>.

Laurel Elder (2004) Why Women Don't Run, *Women & Politics*, 26:2, 27-56, DOI: 10.1300/J014v26n02_02

Leavy, S. (2018). Gender bias in artificial intelligence. *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, 14–16. <https://doi.org/10.1145/3195570.3195580>

Lewis, R., & Anitha, S. (2023). Upskirting: A systematic literature review. *Trauma, Violence, & Abuse*, 24(3), 2003-2018.

Liu, Y. (2024). Unveiling bias in artificial intelligence: Exploring causes and strategies for mitigation. *Applied and Computational Engineering*, 76(1), 124–133. <https://doi.org/10.54254/2755-2721/76/20240576>

Liberia, I., Zurbano, B. y Barredo, D. (2015). "Percepciones de los jóvenes acerca de las actuaciones y discursos públicos sobre la violencia de género en España" *Feminismo/s* 25, 159-182. doi: 10.14198/fem.2015.25.09.

López-Safi, S.B. (2015). La violencia simbólica en la construcción social del género [Symbolic violence in the social construction of gender]. *Academio*, 1(3), pp.30-47.

Marwick, Alice E., & Robyn Caplan. 2018. "Drinking male tears: Language, the manosphere, and networked harassment." *Feminist Media Studies* 18(4): 543-559. DOI: <https://doi.org/10.1080/14680777.2018.1450568>

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), 1-35.

Meyer, David. "Amazon Killed an AI Recruitment System Because It Couldn't Stop the Tool from Discriminating against Women." *Fortune*. *Fortune*, June 8, 2021. <https://fortune.com/2018/10/10/amazon-ai-recruitment-biaswomen-sexist/>.

Misogyny and Misinformation: An analysis of gendered disinformation tactics during the COVID-19 pandemic - EU DisinfoLab. (n.d.). EU DisinfoLab. <https://www.disinfo.eu/publications/misogyny-and-misinformation:-an-analysis-of-gendered-disinformation-tactics-during-the-covid-19-pandemic/>

Muñiz-Rivas, M. y Cuesta-Roldan, J. (2015). "Violencia de género en entornos virtuales [Gender violence in virtual environments]". *Revista del Cisen Tramas/Maepova* 3(2) 63-72. doi:<http://ppct.caicyt.gov.ar/index.php/cisen/index>

Murphy, Gillian, & Emma Flynn. 2022. "Deepfake false memories." *Memory* 30(4): 480-492. DOI: <https://doi.org/10.1080/09658211.2021.1919715>

Nadeem, Ayesha; Abedin, Babak; and Marjanovic, Olivera, "Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies" (2020). *ACIS 2020 Proceedings*. 27. <https://aisel.aisnet.org/acis2020/27>

Nomura, T. (2020). A possibility of inappropriate use of gender studies in human-robot Interaction. *AI & Society*, 35(3), 751–754. <https://doi.org/10.1007/s00146-019-00913-y>

Onta, N. (2007). Recent books on gender and technology. *Gender, Technology and Development*, 11(3), 401–404. <https://doi.org/10.1177/097185240701100306>

ONU: Asamblea General, Convención sobre la eliminación de todas las formas de discriminación contra la mujer, 18 Diciembre 1979, <https://www.refworld.org/es/leg/intinstrument/unga/1979/es/128505> [accedida 20 March 2024]

Organización de Consumidores y Usuarios. (2024, 20 de noviembre). OCU alerta del uso de sesgos de género, raza y edad en la Inteligencia Artificial. Recuperado de <https://www.ocu.org/organizacion/prensa/notas-de-prensa/2024/sesgosinteligenciaartificial201124>

Quandt, Thorsten. 2018. "Dark Participation." *Media and Communication* 6(4): 36-48. DOI: <https://doi.org/10.17645/mac.v6i4.1519>

Rodríguez Martínez, M., y Gaubert, J. (2020). International Women's Day: How can algorithms be sexist? *Euronews*. <https://www.euronews.com/2020/03/08/international-women-s-day-our-algorithms-are-sexist>

Rosser, M^a A., Suriá, R., Suriá Villegas, E., Moya-Mira, C. y García Teruel, E. (2015). Fomento de buenas prácticas para la prevención del ciberacoso sexista en el marco del EEES. Alicante: Universidad de Alicante.

Smith, P. H., & Bamberger, E. T. (2021). Gender inclusivity is not gender neutrality. *Journal of Human Lactation*, 37(3), 441-443.

Stabile, Bonnie, et al. 2019. "Sex, lies, and stereotypes: Gendered implications of fake news for women in politics." *Public Integrity* 21(5): 491-502. DOI: <https://doi.org/10.1080/10999922.2019.1626695>

Stanovsky, G., Smith, N. A., & Zettlemoyer, L. (2019). Evaluating gender bias in machine translation. *arXiv preprint arXiv:1906.00591*.

- Sutko, D. M. (2020). Theorizing femininity in artificial intelligence: a framework for undoing technology's gender troubles. *Cultural Studies*, 34(4), 567–592. <https://doi.org/10.1080/09502386.2019.1671469>
- Sutko, D. M. (2020). Theorizing femininity in artificial intelligence: a framework for undoing technology's gender troubles. *Cultural Studies*, 34(4), 567–592. <https://doi.org/10.1080/09502386.2019.1671469>
- Tabassum, N., & Nayak, B. S. (2021). Gender stereotypes and their impact on women's career progressions from a managerial perspective. *IIM Kozhikode Society & Management Review*, 10(2), 192–208.
- Topkaya Sevinç, E. (2011). Mobbing with a gender perspective: how women perceive, experience and are affected from it? (Master's thesis, Middle East Technical University).
- UNESCO. (2019). First UNESCO recommendations to combat gender bias in applications using artificial intelligence. Retrieved November 27, 2021, from <https://en.unesco.org/news/first-unescorecommendations-combat-gender-bias-applications-using-artificial-intelligence>
- UNESCO. (2020). Artificial intelligence and gender equality. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000374174>
- UNESCO. (s.f.). Módulo 4: Alfabetización mediática e informacional y aprendizaje intercultural. Recuperado de <https://www.unesco.org/mil4teachers/es/module4>
- Vaitla, B., Bosco, C., Alegana, V., y Wouter, E. (2017). Big Data and the Well-Being of Women and Girls Applications on the Social Scientific Frontier. Data2X. https://www.data2x.org/wp-content/uploads/2019/05/Big-Data-and-theWell-Being-of-Women-and-Girls_.pdf
- Wachter-Boettcher, S. (2017). Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech. Norton & Company.
- Wajcman, J. (2010). Feminist theories of technology. *Cambridge Journal of Economics*, 34(1), 143–152. <https://doi.org/10.1093/cje/ben057>
- Wang, E. (2018). Two dangerous visions: What does it really mean for an algorithm to be biased? The Gradient. <https://thegradient.pub/ai-bias/>
- Wellner, G. P. (2020). When AI Is Gender-biased. *HUMANA. MENTE Journal of Philosophical Studies*, 13(37), 127–150.
- Wilfore, Kristina. 2021. A Digital Resilience Toolkit for Women In Politics: Persisting and Fighting Back Against Misogyny and Digital Platforms' Failures. #ShePersisted. Available on <https://she-persisted.org/our-work/supporting-women-leaders/>
- Wojatzki, M. M. (2018). Do women perceive hate differently: Examining the relationship between hate speech, gender, and agreement judgments.
- Women leaders are better at fighting the pandemic. (2020, June 21). CEPR. <https://cepr.org/voxeu/columns/women-leaders-are-better-fighting-pandemic>

World Economic Forum Boston Consulting Group (BCG). (2018). Towards a reskilling revolution: A future of jobs for all. World Economic Forum.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., y Chang, V. (2017). Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. En M. Palmer, R. Hwa y S. Riedel (Eds.), Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 2979-2989). <https://doi.org/10.18653/v1/D17-1323>

Zimmerman, K. (2018). The future of AI may be female, but it isn't feminist. Venture Beat. <https://venturebeat.com/2018/06/28/the-future-of-ai-may-be-female-but-it-isnt-feminist/>

11.2. Anexo. Programa formativo: Gobernanza Ética y Soberanía Tecnológica en el Periodismo Vasco

Título del programa: Alfabetización crítica y capacitación en IA: Hacia un modelo de redacción aumentada y situada en Euskadi.

1. Justificación y fundamentación científica

Este programa formativo no nace de una propuesta genérica de mercado, sino que se deriva directamente de las necesidades críticas identificadas en la presente tesis doctoral. Los resultados obtenidos en el trabajo de campo evidencian tres vacíos que justifican esta intervención:

1. **Déficit de formación institucional (H2):** Solo el 14,1% de los periodistas vascos posee formación reglada, lo que genera una dependencia del autoaprendizaje reactivo.
2. **Vulnerabilidad lingüística y territorial (H3):** Los modelos comerciales presentan sesgos que diluyen la realidad vasca y el euskera.
3. **Vacío de Gobernanza (H5):** La ausencia de protocolos internos deja la ética en manos de la intuición individual de los profesionales.

2. Objetivos del programa

- **Capacidad de auditoría:** Dotar a los periodistas de herramientas para auditar sesgos (de género, territoriales y lingüísticos) en los *outputs* de modelos comerciales.
- **Soberanía editorial:** Establecer criterios para el uso asistencial de la IA que no comprometan la autonomía ni la identidad informativa de proximidad.
- **Gobernanza situada:** Capacitar en el diseño de protocolos internos de trazabilidad y rendición de cuentas específicos para el contexto bilingüe vasco.

3. Estructura curricular: "El modelo de supervisión humana"

El programa se divide en cinco bloques de intervención directa, diseñados para revertir la "cautela informativa" paralizante y convertirla de este modo en una "adopción crítica consciente".

Módulo 1: Fundamentos técnicos y desmitificación del Algoritmo (6h)

- **Del hallazgo a la práctica:** Superar la brecha detectada entre la centralidad discursiva de la IA y el desconocimiento de su lógica interna.
- **Contenido:** Funcionamiento de las arquitecturas *transformer* y limitaciones del aprendizaje supervisado. Análisis de por qué la IA "alucina" más en contextos de baja representatividad de datos (caso del País Vasco).
- **Taller:** Mapas de riesgos algorítmicos aplicados a la información local.

Módulo 2: Herramientas asistenciales y experimentación situada (6h)

- **Del hallazgo a la práctica:** Fomentar el uso de IA en tareas de bajo riesgo (transcripción, etiquetado) para liberar tiempo para la investigación cualitativa.
- **Contenido:** Uso avanzado de LLMs (ChatGPT, Gemini, Claude) mediante *prompt engineering* específico para la recuperación de información local.
- **Práctica:** Comparación de resultados entre lenguas (castellano vs. euskera) para identificar puntos críticos de error y pérdida de matices.

Módulo 3: Ética interseccional y auditoría de sesgos (6h)

- **Del hallazgo a la práctica:** Respuesta directa a la Hipótesis 4 sobre la amplificación de estereotipos de género y territoriales.
- **Contenido:** Metodologías de auditoría de *outputs*. Cómo detectar la invisibilización de mujeres en fuentes de autoridad generadas por IA.
- **Entregable:** Checklist de validación ética antes de la publicación de contenidos asistidos por IA.

Módulo 4: Verificación híbrida en ecosistemas multilingües (6h)

- **Del hallazgo a la práctica:** Respuesta a la Hipótesis 3 sobre la ineficacia de herramientas comerciales para el *fact-checking* en contextos específicos.
- **Contenido:** Herramientas de detección de *deepfakes* y trazabilidad de fuentes. Estrategias de verificación para medios con recursos limitados.
- **Taller:** Simulación de una crisis de desinformación institucional en el País Vasco y protocolos de respuesta rápida.

Módulo 5: Diseño de protocolos y gobernanza interna (6h)

- **Del hallazgo a la práctica:** Respuesta a la Hipótesis 5 sobre la ausencia de mecanismos de rendición de cuentas.
- **Contenido:** Modelos de transparencia: cómo informar a la audiencia del uso de IA (etiquetado). Creación de comités de ética internos.
- **Proyecto final:** Redacción de una "Guía de Estilo Algorítmico" personalizada para cada medio participante.

4. Metodología situada

- **Bilingüismo operativo:** Todos los materiales y talleres se imparten en euskera y castellano, utilizando *datasets* reales del ecosistema vasco (EiTB, Berria, prensa local).

- **Enfoque Human-in-the-loop:** La formación no enseña a dejar que la IA trabaje sola, sino también a cómo debe el periodista supervisar cada paso del proceso editorial.

5. Impacto esperado (Transferencia de Resultados)

Al finalizar el programa, las redacciones participantes habrán pasado de una fase de "cautela informativa" basada en el desconocimiento, a una fase de "apropiación responsable". Se espera que cada medio desarrolle su propio marco de gobernanza, reduciendo la dependencia de criterios externos y reforzando la confianza de la audiencia vasca a través de la transparencia.

6. Público destinatario y perfiles de competencias

El programa está diseñado bajo una lógica de **transversalidad organizativa**, entendiendo que la gobernanza de la IA no es una tarea exclusiva de los departamentos técnicos, sino una responsabilidad compartida que afecta a toda la jerarquía editorial. Se identifican tres niveles de intervención:

- **Nivel Estratégico (dirección y responsables de innovación):** Orientado a directores de medios, jefes de redacción y responsables tecnológicos de grupos como EiTB, Vocento, Grupo Noticias y medios nativos digitales. El enfoque para este perfil se centra en la toma de decisiones sobre la compra o desarrollo de herramientas, la creación de protocolos de transparencia y la gestión de la responsabilidad legal y ética ante la audiencia.
- **Nivel operativo (periodistas de plantilla y colaboradores):** Dirigido a redactores, editores y creadores de contenido que, como revela la investigación, actualmente operan bajo un modelo de autoaprendizaje. Este grupo incluye de este modo perfiles de prensa escrita, radio y televisión, con especial énfasis en aquellos profesionales que trabajan en entornos bilingües y que también requieren asegurar la calidad lingüística y la precisión territorial de los *outputs* algorítmicos.
- **Nivel de soporte y verificación (documentalistas y fact-checkers):** Perfiles encargados de la trazabilidad y la gestión de archivos, fundamentales para implementar los sistemas de verificación híbrida (humano + máquina) que la tesis propone como solución ante la desinformación automatizada.
- **Nivel de transferencia académica:** Estudiantes de Grado en Periodismo y Comunicación Audiovisual de la UPV/EHU y otras instituciones del País Vasco, con el fin de corregir desde la base el déficit formativo detectado y preparar a las futuras promociones para un mercado laboral de naturaleza híbrida.

11.3. Anexo. Marco normativo para una inteligencia artificial ética y responsable en los medios de comunicación del País Vasco

1. Introducción y justificación

La irrupción de sistemas de inteligencia artificial (IA) en los procesos periodísticos —desde la redacción automatizada hasta la verificación, edición, personalización o distribución de contenidos— exige la definición de un marco normativo específico para los medios de comunicación vascos. Dicho marco debe equilibrar la innovación tecnológica con la responsabilidad editorial, la protección de derechos fundamentales y la preservación de la integridad informativa en un entorno culturalmente diverso y lingüísticamente sensible.

Esta propuesta está basada en las directrices de *Basque Artificial Intelligence Center (BAIC)*, el decálogo ético del Colegio Vasco de Periodistas, el Reglamento Europeo de IA, así como las aportaciones empíricas y teóricas derivadas de esta tesis doctoral, pero adaptada a un contexto de aplicación concreta y con criterios operativos.

1. Principios rectores

1. Respeto a los derechos fundamentales y valores democráticos

Toda implementación de IA en el entorno periodístico debe respetar los derechos humanos, la libertad de expresión, el pluralismo informativo y los principios democráticos.

► *Clave práctica:* incluir cláusulas de derechos fundamentales en los contratos con proveedores de IA.

2. Autonomía profesional y control humano significativo

Los sistemas de IA no deben sustituir la responsabilidad editorial ni las decisiones informativas clave.

► *Clave práctica:* toda pieza generada con IA debe ser revisada por un periodista antes de su publicación.

3. Rendición de cuentas y trazabilidad

La cadena de decisiones generadas mediante IA debe ser comprensible, documentada y auditable.

► *Clave práctica:* implementar logs de actividad para cada sistema de IA empleado.

2. Transparencia y explicabilidad

4. Información clara al público

Los medios deben informar si un contenido ha sido generado o modificado por

sistemas de IA.

- *Clave práctica:* utilizar etiquetas visibles como “Contenido generado con IA y revisado por un periodista”.

5. Documentación interna de procesos

Toda redacción debe contar con un protocolo interno que registre qué herramientas se usan, con qué datos y para qué tareas.

- *Clave práctica:* mantener un “registro de uso de IA” actualizado mensualmente.

3. Evaluación de riesgos

6. Clasificación según nivel de riesgo (RIA)

- Sistemas de alto riesgo: automatización de noticias urgentes, moderación de contenidos, perfiles de usuarios.

- Riesgo limitado: para los asistentes de redacción, generación de borradores y subtitulado automático.

- Sin riesgo significativo: herramientas de transcripción, clasificación de archivos.

7. Evaluaciones previas al despliegue (impacto algorítmico)

Todo sistema de IA debe pasar una auditoría interna antes de su uso para detectar posibles efectos discriminatorios, sesgos o errores.

- *Clave práctica:* checklist con criterios como sesgo de entrenamiento, impacto en minorías lingüísticas, manipulación de la agenda.

4. Equidad lingüística y cultural

8. Protección de las lenguas cooficiales

La IA usada debe ser capaz de trabajar en idiomas como el euskera y el castellano de forma justa y proporcional.

- *Clave práctica:* exigir entrenamiento multilingüe o adaptación lingüística antes de adquirir software.

9. Prevención de sesgos socioculturales

Se debe evitar que los algoritmos refuercen estereotipos de género, etnia o clase.

- *Clave práctica:* realizar tests semestrales con ejemplos sensibles (violencia machista, migración, política local).

5. Privacidad, seguridad y protección de datos

10. Cumplimiento del RGPD y consentimiento informado

- Ningún dato personal puede ser usado para entrenar modelos sin que exista una base legal clara.

- Las fuentes periodísticas deben quedar excluidas de este modo de cualquier procesamiento automático.

- *Clave práctica:* anonimizar datasets internos y establecer una política de

retención de datos.

11. Ciberseguridad adaptada al uso de IA

- Actualizar los protocolos de seguridad para contemplar riesgos como inyecciones de prompt, ataques adversariales o fugas de información.
- *Clave práctica:* contratar auditorías externas anuales.

6. Formación, gobernanza y actualización

12. Capacitación continua del personal

- Los periodistas deben recibir formación práctica y crítica sobre IA (sesgos, uso ético, evaluación de resultados).
- *Clave práctica:* incluir talleres bianuales y formación en alfabetización algorítmica.

13. Creación de un comité interno de ética y tecnología

- Conformado por periodistas, tecnólogos y representantes de la dirección, con reuniones trimestrales.
- Funciones: supervisar la implementación, evaluar incidentes, autorizar herramientas nuevas.

14. Revisión periódica de normas internas

- Este marco debe actualizarse al menos una vez al año, integrando cambios legislativos, tecnológicos y sociales.
- *Clave práctica:* asignar esta revisión al comité mencionado.

7. Participación y derechos de la audiencia

15. Canales de reclamación y derecho a la explicación

- Toda audiencia debe poder solicitar explicaciones sobre contenidos generados por IA o rectificaciones.
- *Clave práctica:* incorporar un formulario de consulta en la web del medio.

16. Transparencia proactiva ante incidentes

- Si una pieza generada con IA ha provocado desinformación o daño, se debe explicar públicamente lo ocurrido, sus causas y medidas correctoras.
- *Clave práctica:* una política de comunicación en momentos de crisis con unas guías redactadas previamente.

12. Cláusula final

Este marco puede ser adoptado como base por medios de comunicación del País Vasco que deseen implementar tecnologías de IA de forma ética, segura y respetuosa con su responsabilidad social. Su aplicación requiere voluntad política, coordinación interna y revisión continua.

BIBLIOGRAFÍA

- Adefioye, S. (2024). The adoption of artificial intelligence into journalism practice: Perspectives from the Ghanaian media industry. *Asian Conference on Media & Mass Communication Official Conference Proceedings*, 405–419. <https://doi.org/10.22492/issn.2186-5906.2024.33>
- Adjin-Tettey, T. D., Muringa, T., Danso, S., & Zondi, S. (2024). The role of artificial intelligence in contemporary journalism practice in two African countries. *Journalism and Media*, 5(3), 846–860. <https://doi.org/10.3390/journalmedia5030054>
- Algaba, A., Gancedo, M., y Gualda, E. (2024). Algorithmic bias and gender representation in AI-generated news: A critical perspective. *Journal of Media Ethics*, 39(2), 112–128. <https://doi.org/10.47772/IJRIS.2024.916SCO0020>
- Amponsah, P. N., & Atianashie, A. M. (2024). Navigating the new frontier: A comprehensive review of AI in journalism. *Advances in Journalism and Communication*, 12(1), 1–17. <https://doi.org/10.4236/ajc.2024.121001>
- Arias Jiménez, B., Rodríguez-Hidalgo, C., Mier-Sanmartín, C., & Coronel-Salas, G. (2022). Use of chatbots for news verification. In *Communication and Applied Technologies: Proceedings of ICOMTA 2022* (pp. 133-143). Singapore: Springer Nature Singapore.
- Asociación de Medios de Información [AMI]. (2023). *Libro blanco de la información 2023*. AMI.
- Bailón, O. M. (2024). Dissonance between technological transformation and journalistic ethics: A critical discussion of the impact of artificial intelligence on the media. *adComunica. Revista Científica de Estrategias, Tendencias e Innovación en Comunicación*, (27), 91–114. <https://doi.org/10.6035/adcomunica.8002>
- Bakke, N. A., & Barland, J. (2022). Disruptive Innovations and Paradigm Shifts in Journalism as a Business: From Advertisers First to Readers First and Traditional Operational Models to the AI Factory. *Sage Open*, 12(2). <https://doi.org/10.1177/21582440221094819>
- Bayer, J. (2024). Legal implications of using generative AI in the media. *Information & Communications Technology Law*, 33(3), 310–329. <https://doi.org/10.1080/13600834.2024.2352694>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bitzer, T., Wiener, M., & Cram, W. A. (2023). Algorithmic transparency: Concepts, antecedents, and consequences – A review and research framework. *Communications of the Association for Information Systems*, 52, 344–370. <https://doi.org/10.17705/1CAIS.05214>

- Blázquez Ruiz, F. J. (2022). La paradoja de la transparencia en la IA: Opacidad y explicabilidad. Atribución de responsabilidad. *Revista Internacional De Pensamiento Político*, 17(1), 261–272. <https://doi.org/10.46661/revintpensampolit.7526>
- Bodó, B. (2019). Selling news to audiences – A qualitative inquiry into the emerging logics of algorithmic news personalization in European quality news media. *Digital Journalism*, 7(8), 1054–1075. <https://doi.org/10.1080/21670811.2019.1624185>
- Bontridder, N., & Pouillet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, 3, e18. <https://doi.org/10.1017/dap.2021.20>
- Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011, December). The socialbot network: when bots socialize for fame and money. In *Proceedings of the 27th annual computer security applications conference* (pp. 93-102).
- Carlson, M. (2017). Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New Media & Society*, 20(5), 1755–1772. <https://doi.org/10.1177/1461444817706684>
- Casero- Ripollés, A. , Fabregat, H. D., & Muñoz, L. A. (2023). Percepciones de la ciudadanía española ante la desinformación en tiempos de la COVID-19: efectos y mecanismos de lucha contra las noticias falsas. *Ícono14*, 21(1), 2.
- Cenoz, J. (2008). Achievements and challenges in bilingual and multilingual education in the Basque Country. *AILA Review*, 21, 13–30. <https://doi.org/10.1075/aila.21.03cen>
- Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717718855>
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the national academy of sciences*, 118(9), e2023301118.
- Connor, K. E., & Cali, D. D. (2024). Artificial intelligence in departments of communication: A course proposal. *Explorations in Media Ecology*, 23(2), 175–198. https://doi.org/10.1386/eme_00204_7
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). Sage.
- Cuéllar Rodríguez, A. (2023). Epistemology and ontology in science: The challenge of artificial intelligence. *Anales de la Real Academia Nacional de Farmacia*, 89(3), 379–386. <https://doi.org/10.53519/analesranf.2023.89.03.09>
- Danzon-Chambaud, S. (2021a). A systematic review of automated journalism scholarship: Guidelines and suggestions for future research. *Open Research Europe*, 1, 4. <https://doi.org/10.12688/openreseurope.13096.1>

- Danzon-Chambaud, S., & Cornia, A. (2021b). Changing or reinforcing the “rules of the game”: A field theory perspective on the impacts of automated journalism on media practitioners. *Journalism Practice*, 17(2), 174–188. <https://doi.org/10.1080/17512786.2021.1919179>
- Deuze, M., & Beckett, C. (2022). Imagination, algorithms and news: Developing AI literacy for journalism. *Digital Journalism*, 10(10), 1913–1918. <https://doi.org/10.1080/21670811.2022.2119152>
- Diakopoulos, N. (2015). Algorithmic accountability. *Digital Journalism*, 3(3), 398–415. <https://doi.org/10.1080/21670811.2014.976411>
- Diakopoulos, N. (2017). Enabling accountability of algorithmic media: Transparency as a constructive and critical lens. *Studies in Big Data*, 25–43. https://doi.org/10.1007/978-3-319-59027-1_2
- Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, 5(7), 809–828. <https://doi.org/10.1080/21670811.2016.1208053>
- Dierickx, L., Opdahl, A. L., Khan, S. A., Lindén, C. G., y Rojas, D. C. G. (2024). A data-centric approach for ethical and trustworthy AI in journalism. *Ethics and Information Technology*, 26(4), Artículo e56. <https://doi.org/10.1007/s10676-024-09801-6>
- Dinçer, E. (2024). Hard and soft skills revisited: Journalism education at the dawn of artificial intelligence. *Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(1), 65–78. <https://doi.org/10.30803/adusobed.1462061>
- Dörr, K. N., y Hollnbuchner, K. (2016). Ethical challenges of algorithmic journalism. *Digital Journalism*, 5(4), 404–419. <https://doi.org/10.1080/21670811.2016.1167612>
- Du, Y. R. (2023). Personalization, echo chambers, news literacy, and algorithmic literacy. *Journal of Broadcasting & Electronic Media*, 67(3), 246–273. <https://doi.org/10.1080/08838151.2023.2182787>
- Edelman, Richard (2023): Trust Barometer. <https://www.edelman.com/trust/2023/trust-baromete>
- 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>
- Ermolova, T. V., Savitskaya, N. V., Dedova, O. V., y Guzova, A. V. (2024). The potential of artificial intelligence in preserving linguistic diversity for future generations. *Science for Education Today*, 14(1), 184–198. <https://doi.org/10.15293/2658-6762.2401.09>
- Eustat - Instituto Vasco de Estadística. (2021). *Encuesta de artes e industrias culturales 2021*. https://www.eustat.eus/estadisticas/tema_37/opt_0/tipo_1/temas.html
- Fairfield, J., & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics*, 29(1), 38–51. <https://doi.org/10.1080/08900523.2014.863126>

- Forja-Pena, T., García-Orosa, B., & López-García, X. (2024). The ethical revolution: Challenges and reflections in the face of the integration of artificial intelligence in digital journalism. *Communication & Society*, 37(3), 237–254. <https://doi.org/10.15581/003.37.3.237-254>
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford university press.
- Fu, W. (2023). AI-news personalization system. *IEEE Access*, 11, 85086–85096. <https://doi.org/10.1109/ACCESS.2023.3303479>
- García-Pablos, A., Perez, N., & Cuadros, M. (2024). *EuSQuAD: Automatically translated and aligned SQuAD2.0 for Basque*. arXiv. <https://doi.org/10.48550/arxiv.2404.12177>
- Godler, Y., Reich, Z., & Miller, B. (2020). Social epistemology as a new paradigm for journalism and media studies. *New Media & Society*, 22(2), 213–229. <https://doi.org/10.1177/1461444819856922>
- Google. (2024). The ROI of generative AI. https://services.google.com/fh/files/misc/the_roi_of_generative_ai.pdf
- Gutiérrez-Caneda, B., & Vázquez-Herrero, J. (2024). Redibujando las líneas contra la desinformación: Cómo la IA está modificando el presente y futuro del fact-checking. *Tripodos*, (55), 71–89. <https://doi.org/10.51698/tripodos.2024.55.04>
- Gutiérrez-Caneda, B., Vázquez-Herrero, J., & López-García, X. (2023). AI application in journalism: ChatGPT and the uses and risks of an emergent technology. *El Profesional de la Información*, 32(2), e320203. <https://doi.org/10.3145/epi.2023.mar.03>
- Guzman, A. L., & Lewis, S. C. (2024). What generative AI means for the media industries, and why it matters to study the collective consequences for advertising, journalism, and public relations. *Emerging Media*, 2(3), 347–355. <https://doi.org/10.1177/27523543241289239>
- Helberger, N., Diakopoulos, N., Casari, G., Kuipers, M., Trilling, D., & Vrijenhoek, S. (2022). Towards a normative perspective on journalistic AI. *Digital Journalism*, 10(10), 1605–1626. <https://doi.org/10.1080/21670811.2022.2152195>
- Hine, C. (2015). *Ethnography for the internet: Embedded, embodied and everyday*. Routledge. <https://doi.org/10.4324/9781003085348>
- Ioscote, F., Gonçalves, A., & Quadros, C. (2024). Artificial intelligence in journalism: A ten-year retrospective of scientific articles (2014–2023). *Journalism and Media*, 5(3), 873–891. <https://doi.org/10.3390/journalmedia5030056>
- Jandrić, P. (2019). The postdigital challenge of critical media literacy. *The International Journal of Critical Media Literacy*, 1(1), 26–37. <https://doi.org/10.1163/25900110-00101002>
- Jones, B., Jones, R., & Luger, E. (2022). AI ‘everywhere and nowhere’. *Digital Journalism*, 10(10), 1731–1755. <https://doi.org/10.1080/21670811.2022.2145328>

- Kanasheva, S. J., Duisengazy, S., & Seitzhanova, Z. (2024). Artificial intelligence in journalism: Foreign experience. *Bulletin of the L. N. Gumilyov Eurasian National University. Journalism Series*, 146(1), 45–54. <https://doi.org/10.32523/2616-7174-2024-3-148-7-19>
- Karlsson, M. (2021). *Transparency and journalism*. Routledge. <https://doi.org/10.4324/9780429352935>
- Kasperuniene, J. (2021). The use of artificial intelligence in social research: Multidisciplinary challenges. En *Advances in Intelligent Systems and Computing* (pp. 312–324). Springer. https://doi.org/10.1007/978-3-030-70187-1_22
- Kim, H. (2019). *AI in journalism: Creating an ethical framework* [Tesis de honor, Syracuse University]. SURFACE. https://surface.syr.edu/honors_capstone/1083
- Kioko, P. M., Booker, N., Chege, N., & Kimweli, P. (2022). The adoption of artificial intelligence in newsrooms in Kenya: A multi-case study. *European Scientific Journal, ESJ*, 18(22), 278–294. <https://doi.org/10.19044/esj.2022.v18n22p278>
- Kohring, M., & Matthes, J. (2007). Trust in news media: Development and validation of a multidimensional scale. *Communication research*, 34(2), 231–252.
- Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. Sage Publications.
- Latzer, M., & Just, N. (2020). Governance by and of algorithms on the internet: Impact and consequences. *Oxford Research Encyclopedia of Communication*. <https://doi.org/10.1093/acrefore/9780190228613.013.904>
- Lewis, S. C., & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society*, 35(5), 602–619. <https://doi.org/10.1177/0163443713485494>
- Lewis, S. C., Guzman, A. L., & 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>
- Lindén, C. G. (2017). Algorithms for journalism: The state of the art and its implications. *The Journal of Media Innovations*, 4(1), 60–76. <https://doi.org/10.5617/jmi.v4i1.2420>
- Liu, S., Wang, Y., y Zhang, L. (2024). Hidden biases: Analyzing gender stereotypes in large language models. *Information Processing & Management*, 61(3), Artículo 103682. <https://doi.org/10.1016/j.ipm.2024.103682>
- Lopez, M. G., Porlezza, C., & MacFarlane, A. (2022). A question of design: Strategies for embedding AI-driven tools into journalistic work routines. *Digital Journalism*, 10(7), 922–939. <https://doi.org/10.1080/21670811.2022.2043759>
- Lopezosa, C., Codina, L., Pont-Sorribes, C., & Váñez, M. (2023). Use of generative artificial intelligence in the training of journalists: Challenges, uses and training proposal. *El Profesional de la Información*, 32(4), e320408. <https://doi.org/10.3145/epi.2023.jul.08>

- Makeleni, S., Mutongoza, B., y Linake, M. (2023). Language education and artificial intelligence: An exploration of challenges confronting academics in Global South universities. *Journal of Culture and Values in Education*, 6(2), 158–171. <https://doi.org/10.46303/jcve.2023.14>
- Masís-González, T. M. (2024). La metáfora del shoggoth en la inteligencia artificial. *Pensamiento Actual*, 24(43), 89-95.
- Mohamed, S., Png, M., y Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4), 659–684. <https://doi.org/10.1007/s13347-020-00405-8>
- Montal, T., y Reich, Z. (2016). I, Robot. You, Journalist. Who is the author? *Digital Journalism*, 5(7), 829–849. <https://doi.org/10.1080/21670811.2016.1209083>
- Monti, M. (2019). Automated journalism and freedom of information. *Opinio Juris in Comparatione*, 1, 1–18.
- Montoro-Montarroso, A., Cantón-Correa, J., Rosso, P., Chulvi, B., Panizo-Lledot, Á., Huertas-Tato, J., Calvo-Figueras, B., Rementería, M. J., y Gómez-Romero, J. (2023). Fighting disinformation with artificial intelligence: Fundamentals, advances and challenges. *El Profesional de la Información*, 32(3), e320322. <https://doi.org/10.3145/epi.2023.may.22>
- Negreira-Rey, M. C., Vázquez-Herrero, J., y López-García, X. (2023). No people, no news: News deserts and areas at risk in Spain. *Media and Communication*, 11(3), 235–246. <https://doi.org/10.17645/mac.v11i3.6727>
- Newman, N., Fletcher, R., Robertson, C. T., Arguedas, A. R., & Nielsen, R. K. (2024). Digital News Report 2024. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024>
- Newsguard. (2023). *Unreliable AI-generated news and information sites*. <https://www.newsguardtech.com/special-reports/ai-tracking-center/>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., y Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Artículo 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., y Qiao, M. S. (2021a). AI literacy: Definition and ethical issues. *Proceedings of the Association for Information Science and Technology*, 58(1), 504–509. <https://doi.org/10.1002/pra2.487>
- Novelli, C., Casolari, F., Rotolo, A., Taddeo, M., & Floridi, L. (2024). Taking AI risks seriously: a new assessment model for the AI Act. *Ai & Society*, 39(5), 2493-2497.
- Nurlatifah, M., e Irwansyah, I. (2019). Fact-checking journalism sebagai platform kolaborasi human and machine pada jurnalisme digital. *Jurnal Komunikasi*, 13(2), 121–134. <https://doi.org/10.20885/komunikasi.vol13.iss2.art1>
- Observatorio Vasco de la Cultura. (2023). *Informe del sector de medios de comunicación y publicidad en Euskadi 2023*. Gobierno Vasco.

- Palomo, B., & Sedano, J. (2021). Cross-media alliances to stop disinformation: A real solution? *Media and Communication*, 9(1), 239-250. <https://doi.org/10.17645/mac.v9i1.3535>
- Palomo, B., Tandoc Jr, E. C., & 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://dx.doi.org/10.5209/esmp.92160>
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. penguin UK.
- Pavlik, J. V. (2014). Transformation: Examining the implications of emerging technology for journalism, media and society. *Athens Journal of Mass Media and Communications*, 1(1), 9–24. <https://doi.org/10.30958/ajmmc.1-1-1>
- Peña-Fernández, S., Meso-Ayerdi, K., Larrondo-Ureta, A., y Díaz-Noci, J. (2023). Without journalists, there is no journalism: The social dimension of generative artificial intelligence in the media. *El Profesional de la Información*, 32(2), e320227. <https://doi.org/10.3145/epi.2023.mar.27>
- Pérez-Dasilva, J., Peña-Alonso, U., Ganzabal-Learreta, M., Meso-Ayerdi, K., & Mendiguren-Galdospin, T. (2026). Percepciones sobre el uso de IA en el periodismo vasco y su integración en los medios. *Revista Latina de Comunicación Social*, (84), 1-28. <https://doi.org/10.4185/rlcs-2026-2517>
- Perez, Luciano & Perez, Rubén & Seca, Maria. (2023). Metodología de la investigación científica.
- Peterson-Salahuddin, C., y Diakopoulos, N. (2020). Negotiated autonomy: The role of social media algorithms in editorial decision making. *Media and Communication*, 8(3), 27–38. <https://doi.org/10.17645/mac.v8i3.3001>
- Pinto, M. C., y Barbosa, S. O. (2024). Artificial intelligence in Brazilian digital journalism: Historical context and innovative processes. *Journalism and Media*, 5(1), 122–138.
- Porlezza, C. (2023). Promoting responsible AI: A European perspective. *Communications*, 48(3), 370–394.
- Prem, E. (2023). From ethical AI frameworks to tools. *AI and Ethics*, 3(3), 699–716.
- Rossetti, M., & Zaman, T. (2023). Bots, disinformation, and the first impeachment of US President Donald Trump. *Plos one*, 18(5), e0283971.
- Salas-Pilco, S. Z., Xiao, K., y Oshima, J. (2022). Artificial intelligence and new technologies in inclusive education for minority students: A systematic review. *Sustainability*, 14(20), Artículo 13572. <https://doi.org/10.3390/su142013572>
- Salgado, L. (2022). From the classroom to the newsroom: A critical route to introduce AI in journalism education. *Facts & Frictions*, 2(1), 80–97. <https://doi.org/10.22215/ff/v2.i1.04>
- Sánchez-García, P., Merayo Álvarez, N., Calvo Barbero, C., & Díez-Gracia, A. (2023). Spanish technological development of artificial intelligence applied to journalism: companies and tools for documentation, production and distribution of information. *El Profesional de la información*, 32(2).

- Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2023). Artificial Intelligence Governance For Businesses. *Information Systems Management*, 40(3), 229–249. <https://doi.org/10.1080/10580530.2022.2085825>
- Showkat, D. (2022). Supporting responsible data practices in the news media. *Proceedings of the 6th HUMANIZE Workshop*.
- Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, 10(10), 1832–1854. <https://doi.org/10.1080/21670811.2022.2063150>
- Sonni, A. F., Hafied, H., Irwanto, I., y Latuheru, R. (2024). Digital newsroom transformation: A systematic review of the impact of artificial intelligence on journalistic practices, news narratives, and ethical challenges. *Journalism and Media*, 5(4), 1554–1570. <https://doi.org/10.3390/journalmedia5040097>
- Strömbäck, J., Tsifti, Y., Boomgaarden, H., Damstra, A., Lindgren, E., Vliegenthart, R., & Lindholm, T. (2020). News media trust and its impact on media use: Toward a framework for future research. *Annals of the International Communication Association*, 44(2), 139-156.
- Ta, R., y Lee, N. (2023). How language gaps constrain generative AI development. *International Journal of Comparative Studies in International Relations and Development*, 9(1), 48–52. <https://doi.org/10.48028/iiprds/ijcsird/ijcsird.v9.i1.03>
- Taylor, L. (2017). What is data justice? *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717736335>
- Tejedor, S., y Vilà, P. (2021). Exo journalism: A conceptual approach to a hybrid formula between journalism and artificial intelligence. *Journalism and Media*, 2(4), 830–840. <https://doi.org/10.3390/journalmedia2040048>
- Thurman, N., Dörr, K., y Kunert, J. (2017). When reporters get hands-on with robo-writing. *Digital Journalism*, 5(10), 1240–1259. <https://doi.org/10.1080/21670811.2017.1289819>
- Tomassi, A., Falegnami, A., y Romano, E. (2024). Mapping automatic social media information disorder. The role of bots and AI in spreading misleading information in society. *PLOS ONE*, 19(5), Artículo e0303183. <https://doi.org/10.1371/journal.pone.0303183>
- Trillo-Domínguez, M., & Alberich-Pascual, J. (2020). Análisis y tipificación de formatos emergentes en el ciberperiodismo español: de la adaptación multimedia a la disrupción transmedia. *Estudios sobre el Mensaje Periodístico*, 26(1), 367-375. <https://dx.doi.org/10.5209/esmp.67317>
- Tsekhmeistruk, R. (2024). Quantifying algorithmic bias in news recommendations: Methodologies and case studies. *Scientific Journal of Polonia University*, 66(5), 251–259.
- Túñez-López, J. M., Fieiras Ceide, C., & Vaz-Álvarez, M. (2021). Impacto de la Inteligencia Artificial en el Periodismo: transformaciones en la empresa, los productos, los contenidos y el perfil profesional. *Communication & Society*, 34(1), 177-193.

- Ufarte Ruiz M. J., Calvo Rubio L. M. y Murcia Verdú F. J. (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>
- Vázquez, J. R. L., Paz, D. G., Vega, E. L., Quevedo, L. D., & Guerra, Y. H. (2016). Del Gatekeeper al Content Curator: cambiar algo para que no cambie nada. *Razón y palabra*, (92), 44-23.
- Vraga, E. K., Bode, L., & Tully, M. (2022). Creating news literacy messages to enhance expert corrections of misinformation on Twitter. *Communication Research*, 49(2), 245-267.
- Yeung, W. N., & Dodds, T. (2024). *Automated journalism*. arXiv. <https://doi.org/10.48550/arXiv.2409.03462>
- Zheng, Y., Mokander, J., y Floridi, L. (2025). Human-in-the-loop strategies for algorithmic auditing in newsrooms. *Nature Machine Intelligence*. [En prensa].