

RAE-IC, Revista de la Asociación Española de Investigación
de la Comunicación

vol. 11, núm. Especial (2024), raeic11e03

ISSN 2341-2690

DOI: <https://doi.org/10.24137/raeic.11.e.3>

Recibido el 9 de mayo de 2024

Aceptado el 11 de julio de 2024



Verificación automática bajo vigilancia humana: implementación de IA en el caso de Newtral

*Automated fact-checking under human surveillance: AI implementation in
the case of Newtral*

Quiles-Morán, Borja

Universidad Miguel Hernández de Elche (UMH)

fquiles@umh.es

Arias Robles, Félix

Universidad Miguel Hernández de Elche (UMH)

farías@umh.es

Forma de citar este artículo:

Quiles-Morán, B. & Arias-Robles, F. (2024). Verificación automática bajo vigilancia humana: implementación de IA en el caso Newtral. *RAE-IC, Revista de la Asociación Española de Investigación de la Comunicación*, 11(Especial), raeic11e03. <https://doi.org/10.24137/raeic.11.e.3>

Resumen:

La Inteligencia Artificial (AI), introducida de forma gradual en las redacciones periodísticas desde comienzos de la década de 2010, puede ayudar a luchar contra la

desinformación en un contexto actual de crisis de confianza en el periodismo tradicional y proliferación de bulos. Este trabajo analiza las herramientas de IA implementadas en Newtral, un medio especializado en verificación de referencia en España, para agilizar el flujo de trabajo en la redacción y acelerar el proceso de publicación de desmentidos. Mediante entrevistas semiestructuradas a cuatro profesionales (tres del equipo de tecnología y una redactora), se constata que los proyectos que se apoyan en IA facilitan la labor de los periodistas y generan discusiones y sinergias positivas entre ingenieros y profesionales de la información. La IA se emplea para detectar temas relevantes y hechos verificables, transcribir de audio a texto y monitorear mensajes en redes sociales. Para evitar errores de los algoritmos, se emplean técnicas human-in-the-loop que aseguren que nada que haya sido manipulado por la IA se publique sin revisión humana.

Palabras clave: inteligencia artificial, periodismo, Newtral, verificación, estudio de caso.

Abstract:

Artificial Intelligence (AI), which has been progressively introduced into newsrooms since the early years of the 2010s, can serve as a potent tool to combat disinformation in the current scenario of a crisis of trust in traditional journalism and the rampant spread of hoaxes. This research paper delves into the AI tools that have been implemented in Newtral, a leading verification media outlet in Spain, with the aim to streamline the workflow in the newsroom and expedite the process of publishing disinformation. Through semi-structured interviews conducted with four professionals (comprising three from the technology team and one editor), it is discovered that AI projects significantly facilitate the work of journalists and foster positive discussions and synergies between engineers and information professionals. AI is employed to detect relevant topics and verifiable facts, transcribe audio content to text, and monitor social media posts. To circumvent algorithmic errors, human-in-the-loop techniques are employed to ensure that no content that has been manipulated by AI is published without undergoing human review.

Keywords: artificial intelligence, journalism, Newtral, fact-checking, case study

1. INTRODUCCIÓN

La Inteligencia Artificial (IA) es ya uno de los principales motores de cambio en la sociedad actual y llega más allá que cualquier otro avance computacional anterior (Cobisa, 2022). El periodismo no ha permanecido ajeno a este fenómeno: la profesión se encuentra en un contexto tecnológico propicio para la experimentación con herramientas relacionadas con el procesamiento y la generación automática del lenguaje, los Large Language Models (LLM) y el *machine learning*. Estos avances previsiblemente supondrán un paso adelante en toda la cadena de innovaciones que se han dado desde que comenzó el proceso de digitalización (Beckett & Yaseen, 2023; Broussard et al., 2019; López-García & Vizoso, 2021; Túnñez-López & Tejedor-Calvo, 2019). La crisis post-pandemia, además, ha agudizado la necesidad de los medios de comunicación por innovar para no quedarse rezagados en un entorno complejo (García-Avilés et al., 2022).

Se estima que casi la mitad del proceso de adopción de esta tecnología en los medios de comunicación se producirá a corto plazo y que las mejoras en la productividad laboral y la automatización de tareas serán clave (Rao & Verweij, 2022). Las consecuencias de esta transformación demandan una reflexión crítica en los profesionales y las audiencias sobre temas como la responsabilidad, la ética o la independencia de una herramienta que, cada vez más, se concentra en un pequeño conjunto de actores del sector privado (Maslej et al., 2023; Simon, 2022).

La investigación sobre la aplicación de IA al periodismo, pese a ser relativamente nueva, se está consolidando y generando una producción académica madura (García-Orosa et al., 2023; Parratt-Fernández et al., 2021). Dentro de la academia, el interés por el impacto de la IA en el periodismo ha crecido de forma gradual dada la “relativa novedad de la disciplina”, lo que justifica el enfoque exploratorio y descriptivo de herramientas, metodologías y aplicaciones que toman estos estudios (Calvo-Rubio & Ufarte-Ruiz,

2021, p. 167). Los temas más estudiados son la aplicación al periodismo de datos, la redacción por robots y la verificación de noticias (Parratt-Fernández et al., 2021), mientras que otros asuntos como la responsabilidad de los algoritmos, el valor simbólico del periodismo automatizado y las cuestiones éticas que este fenómeno entraña permanecen, todavía, poco explorados (Ananny, 2016; Diakopoulos, 2015; Dörr & Hollnbuchner, 2017; Lewis & Zamith, 2017; Thurman et al., 2019).

En los medios especializados en *fact-checking* (en adelante, verificación), la IA se ha convertido en un aliado importante para comprobar la veracidad de una información (García-Marín, 2022, p. 8), aunque el fenómeno solo ha sido estudiado de forma limitada por la academia (Johnson, 2023, p. 2).

Este artículo busca ampliar el conocimiento en el área de la aplicación de IA para la verificación de contenidos con el estudio de caso de Newtral. Newtral es una *start-up* fundada en 2018 por periodistas del programa de televisión “El Objetivo” liderados por Ana Pastor, propietaria del medio. El medio se centra, principalmente, en temas políticos (García-Vivero & López, 2021). Forma parte, junto a Maldita, EFE Verifica y Verificat, de la International Fact-Checking Network (IFCN), asociación que obliga a aceptar y cumplir un código de principios éticos (Aguado-Guadalupe & Bernaola-Serrano, 2020).

Pese a su corta trayectoria, Newtral ha despertado el interés de la academia para estudiar asuntos como la verificación de contenido en la pandemia de covid-19 (Almansa-Martínez et al., 2022), la desinformación en el primer año de guerra entre Rusia y Ucrania (Magallón-Rosa et al., 2023) o la desacreditación de bulos relacionados con el feminismo (Meléndez-Malavé & Pérez-Quintana, 2023). En los últimos años se han producido, además, artículos científicos relacionados con herramientas de IA desarrolladas por Newtral, que incluyen experimentos hechos por el medio (Beltrán et al., 2021; Larraz et al., 2023) o comparativas de los proyectos de Newtral con los de otros medios (Mondría-Terol, 2023).

Este estudio se ha realizado a partir de entrevistas semiestructuradas con cuatro profesionales de Newtral, tres del equipo de tecnología y uno de la parte periodística. El principal objetivo de la investigación es clasificar las herramientas de IA que se aplican en Newtral y analizar cómo ayudan a acelerar y profundizar en la verificación de desinformación. Los objetivos específicos que derivan de este son comprobar la relación entre los perfiles periodísticos y técnicos que trabajan con herramientas de IA y conocer los problemas éticos que puedan surgir de la aplicación de IA. Se han planteado las siguientes preguntas de investigación:

- RQ1: ¿Cuáles son los proyectos de IA implementados, en desarrollo o descartados y cuáles son sus características?
- RQ2: ¿Cómo aplican los periodistas esta tecnología en el desempeño de su trabajo?
- RQ3: ¿Cómo es la colaboración entre periodistas e ingenieros para el desarrollo de proyectos conjuntos?
- RQ4: ¿Qué trabas éticas se encuentran en el desarrollo de herramientas basadas en IA?

2. MARCO TEÓRICO

2.1. LA IA APLICADA AL PERIODISMO

La publicación a comienzos de la década de 2010 de las primeras noticias creadas automáticamente en medios de referencia como *Los Angeles Times* (Oremus, 2014) inició una carrera para introducir la IA en el periodismo. Los siguientes años comenzó a implementarse favorablemente en medios estadounidenses, chinos, japoneses o británicos entre otros (Ufarte-Ruiz & Manfredi-Sánchez, 2019). El mercado periodístico español no fue ajeno a este fenómeno y destacan ejemplos como el proyecto Medusa de Vocento, el acuerdo entre EFE y RTVE con la compañía Narrativa (Túñez-López et al., 2021) o la IA de *El Confidencial* para la monetización de su contenido y la automatización de la información deportiva (Rojas-Torrijos & Toural-Bran, 2019).

La IA se implementa casi siempre como una herramienta complementaria con la labor humana que permite a los periodistas centrarse en tareas más creativas y que aporten valor y, además, establece una relación integradora entre humanos e IA, ya que el contenido creado por ambos puede ser corregido por el otro (Ufarte-Ruiz & Manfredi-Sánchez, 2019; van Dalen, 2012; Wu et al., 2019).

El uso más conocido y llamativo de la IA es la generación de contenido, hasta hace poco limitada a información económica, deportiva o meteorológica (Túñez-López et al., 2021). Sin embargo, estas herramientas se expanden para facilitar actividades específicas en el proceso de producción de contenido, como “la búsqueda de mayor objetividad para evitar errores y ambigüedades, la detección de temas relevantes para construir la agenda, la recopilación de información, la traducción automática y el desarrollo de sistemas de recomendación de noticias” (Canavilhas, 2022, p. 3). Otros autores destacan el rol transformador de la IA en las etapas de captación y de distribución de noticias y de la redacción de textos (Wu et al., 2019), así como para facilitar una personalización de la oferta de contenidos y el diseño de los muros de pagos a través de algoritmos (Carvajal et al., 2022).

Los contenidos automatizados, pese a ser menos legibles que los producidos por humanos, han demostrado ser igual de creíbles y de la misma calidad ante los ojos de la audiencia (Haim & Graefe, 2017), una percepción positiva que trasciende incluso fronteras culturales (Zheng et al., 2018). La automatización, no obstante, supone problemas éticos y sociales que precisarán de nuevas habilidades en la audiencia (Thurman et al., 2017). En las redacciones, además, amplios sectores ven la IA como un peligro y no como un complemento (Mondría-Terol, 2023), temores infundados según la mayoría de las investigaciones (Ufarte-Ruiz & Manfredi-Sánchez, 2019), aunque algunos estudios evidencien el uso de la automatización como estrategia para reducir costes humanos (Kim & Kim, 2017).

Para hacer frente a estos desafíos que la IA plantea, son necesarios los cambios en la cultura organizacional, ya que solo aquellas compañías capaces de fomentar una cultura de aprendizaje alrededor de esta tecnología podrán transformarse y sobrevivir a un

contexto complejo (Cameron & Quinn, 2011; Porcu et al., 2020). A medida que la IA se adapte más al periodismo, habrá mayor interacción entre profesionales de la información y máquinas, lo que supondrá una reformulación de la organización del trabajo en los medios (Parratt-Fernández et al., 2021). Otros obstáculos para el desarrollo de la IA en los medios de comunicación son la falta de profesionales con formación y habilidades tecnológicas (Mondría-Terol, 2023) o la dificultad de los algoritmos para codificar computacionalmente aquello noticioso (Stray, 2019).

2.2 VERIFICACIÓN E INNOVACIÓN TECNOLÓGICA

La desinformación se ha erigido en la última década no solo como una amenaza a la seguridad, sino también a las instituciones y los procesos democráticos (Bennett & Livingston, 2018; Waisbord, 2018). Los elementos que han contribuido a potenciar los efectos de la desinformación son la crisis de los medios, la precarización de los periodistas, los nuevos medios digitales, la explosión de las redes sociales (Rojas-Caja, 2020) y la crisis sanitaria causada por la covid-19 en 2020 (Pérez-Escoda et al., 2021; Zarocostas, 2020).

Para luchar contra esta “pandemia” de desinformación ya se habían creado anteriormente herramientas como los medios de verificación, que, aunque nacieron para verificar aquello publicado por otros medios, en la última década han ampliado el foco al análisis a mensajes políticos, científicos y económicos difundidos, sobre todo, en la red (Cazalens et al., 2018). Los verificadores se enfrentan a una desinformación que se difunde de forma mucho más rápida que los verificados, lo que los obliga a ser efectivos y proactivos (Vosoughi et al., 2018).

La tecnología se convierte así no solo en un aliado, sino en la base de los verificadores a la hora de cotejar información y difundirla (Vázquez-Herrero et al., 2019). Las posibilidades de la automatización ofrecen un “potencial prometedor” que los verificadores ya explotan pese a estar aún en una fase temprana (García-Marín, 2022, p. 15). La verificación automatizada, precisa y en directo se ha convertido en el horizonte de los medios de verificación y en el objetivo de sus innovaciones (Hassan et al., 2015).

Para conseguirla, se necesitan cambios computacionales fundamentales, que han despertado el interés de los científicos de la comunicación, tales como la mejora de los sistemas de Procesamiento de Lenguaje Natural (NLP por sus siglas en inglés) y el aumento de la capacidad de recolección automática de la información y el desarrollo de herramientas capaces de enfrentar las complejidades lingüísticas y de escasez de información sobre los bulos (Cazalens et al., 2018; de Oliveira et al., 2021).

A nivel internacional, pueden nombrarse medios como Full Fact, PolitiFact, FactCheck.org, Washington Post (Graves, 2018), Logically (López-Marcos y Vicente-Fernández, 2021), FactStream o el proyecto News Tracer de Reuters (Beckett & Yaseen, 2023), que emplean herramientas de IA para labores de verificación.

En el caso de los medios especializados en verificación dentro de España, Newtral destaca por su desarrollo en el campo de la tecnología, específicamente en los proyectos de IA, como demuestran estudios anteriores como el de Mondría-Terol (2023, pp. 50-54). El otro gran medio de verificación español por volumen de seguidores y producción de contenido, Maldita.es, posee herramientas de IA como un chatbot con el que los usuarios pueden interactuar (*Maldita.es lanza el primer chatbot de Whatsapp en español para verificar bulos automáticamente*, 2020).

El modelo adoptado por la mayoría de los verificadores que usan IA a la hora de verificar es la técnica *human-in-the-loop*, un sistema que permite la participación humana en los procesos automáticos (Wu et al., 2021) y mejora la credibilidad y la transparencia de los productos finales, además de aumentar la eficiencia y la efectividad de la verificación (Demartini et al., 2020) en un mundo en el que las personas son necesarias para la atribución y la exactitud de la información y en el que los periodistas desean mantener su autoridad y su autonomía frente a los algoritmos (Johnson, 2023).

3. METODOLOGÍA

Esta investigación profundiza en el desarrollo y uso de las herramientas de IA integradas en el medio de verificación Newtral. El caso se ha elegido principalmente por disponer de un equipo de tecnología, lo que ha permitido comparar las distintas visiones tienen

técnicos y periodistas sobre la aplicación de IA en la verificación, y por la integración, en el momento de la investigación, de varios proyectos con IA en el flujo de trabajo de la redacción (Beltrán et al., 2021).

Se ha elegido para realizar este trabajo una metodología cualitativa basada en entrevistas semiestructuradas (Valles, 2002). A lo largo de los meses de abril y mayo de 2022, se realizaron 4 entrevistas a miembros del equipo de Newtral (3 del equipo tecnológico y 1 de la redacción) (Tabla 1). Para concertar las entrevistas, se usó la técnica de muestreo de la bola de nieve. Según este modelo, a través de un individuo elegido de una muestra finita (en este caso, E1), se consigue el contacto de otros individuos de la misma muestra (Goodman, 1961).

Tabla 1. Datos sobre los entrevistados

N	Edad	Sexo	Formación	Otras formaciones	Puesto de trabajo	Horas dedicadas a la IA (%)
E1	43	H	Ingeniería de Telecomunicaciones	Doctor en Ingeniería Informática	Responsable de tecnología	30%
E2	33	M	Periodismo	Doctoranda en Periodismo	Coordinadora de Verificación	N/C
E3	22	M	Ingeniería Matemática en Ciencia de Datos		ML Engineer	80%
E4	41	H	Ingeniería Electrónica	Cursos de programación	Tech lead	25%

Fuente: Elaboración propia.

Se realizaron entrevistas a fondo con un cuestionario semiestructurado a través de una aplicación de videoconferencias. Este cuestionario contenía entre 10 y 15 preguntas diferentes para cada uno y centradas en su especialidad. Los bloques temáticos fueron la descripción de su labor en Newtral y del equipo en el que trabajan, su relación con el resto de los equipos, las herramientas de IA utilizadas o desarrolladas en su trabajo

diario, la utilidad de estas IA en su trabajo, los proyectos y necesidades de IA de cara al futuro y los límites éticos de estas IA en el contexto de su medio.

4. RESULTADOS

4.1. ASPECTOS GENERALES DE LA IA EN NEWTRAL

Los proyectos de Newtral se basan en una IA de tipo NLP (E1), un campo donde se mezclan disciplinas como el *big data* o la lingüística. A diferencia de otros medios que utilizan métodos de generación de lenguaje (NLG), Newtral tan solo emplea la IA como herramienta complementaria de las actividades humanas y siempre mediante la técnica *human-in-the-loop*, por la que humano y máquina se retroalimentan en todo momento para que el resultado sea mejor y más fiable. En ningún momento han trabajado en “sistemas puramente automatizados salvo en aquellos escenarios donde es algo muy claro y ya casi podría decirse que no es IA” (E1).

En la misma línea, E2 coincide con E1 en que la IA no es del todo fiable y no cree, desde su experiencia trabajando en el día a día con esta tecnología, que un medio de verificación pueda verificar sin intermediación humana, ya que un robot “no es capaz de detectar todos los elementos contextuales” (E2).

El trabajo en Newtral se divide en tres equipos: tecnología, verificación del discurso político y bulos. El departamento de tecnología está formado por entre 5 y 8 personas, entre las que se incluyen E1, E3 y E4. Es un departamento muy inestable, sobre todo, porque los especialistas que trabajan con IA tienen un perfil muy demandado por empresas que ofrecen salarios y condiciones contra las que Newtral es incapaz de competir (E1). La parte de verificación puede dividirse en otros dos equipos: el de verificación del discurso político y desacreditación de bulos. En la parte de verificación del discurso político trabajan 7 mujeres periodistas con distintos perfiles (E2). Su función es monitorizar las intervenciones y acciones políticas para fiscalizarlas y contrastarlas. El equipo de desacreditación, por su parte, es el encargado de invalidar bulos y difundir su verificación.

Estos equipos descritos, pese a pertenecer a disciplinas diferentes, han evolucionado hasta tener una relación cercana y directa por la inclusión de la tecnología en las redacciones. Los cuatro entrevistados cuentan que la relación entre periodistas y técnicos es muy buena; se mantienen reuniones habituales, se trabaja de forma conjunta en proyectos y tratan de tener una comunicación fluida, algo sencillo al compartir el mismo espacio de trabajo y las zonas comunes. No obstante, tanto periodistas como técnicos destacan el choque de mentalidades. Para E2 y E4, los periodistas tienden a ser más creativos y caóticos, frente a unos técnicos mucho más meticulosos. Esto provoca que ambas visiones no siempre coincidan, sobre todo en la aplicación de tecnología en áreas que los periodistas no consideran necesaria. Estos debates, reconocen ambos grupos, son beneficiosos y siempre acaba primando la visión más ética y deontológica (E4).

4.2. PROYECTOS DE IA EN NEWTRAL

En Newtral, el flujo de trabajo de la redacción se encuentra dividido en cuatro procesos: 1) monitorización, 2) detección de hechos relevantes, 3) verificación del dato y 4) publicación. La IA está completamente implementada en los dos primeros procesos, mientras que los intentos de usarla en la parte de verificación son mínimos y están algo estancados (E1). Esta parte es la que más interesa automatizar de cara al futuro porque es la que genera el “cuello de botella” del flujo de trabajo (E1), aunque en el presente resulte imposible de realizar asegurando unos principios éticos y una calidad periodística válida.

Newtral tenía, en el momento de la investigación, cinco proyectos de Inteligencia Artificial en activo. Además, otro proyecto se encuentra paralizado pese a su alto grado de desarrollo. Para E1, el rol de la IA en un medio de verificación es “hacer que sea más rápido el proceso; poder analizar volúmenes de datos más grandes con el mismo personal y de forma más rápida”. Además de estos proyectos principales, pensados para su uso generalizado por toda la redacción, la Inteligencia Artificial se aplica en proyectos periodísticos concretos (reportajes o investigaciones).

Tabla 2. Proyectos de IA desarrollados en Newtral

N	Nombre	Descripción	Estado
P1	Transcripción de audio a texto	Herramienta que transcribe automáticamente textos.	Implementado
P2	Claim detection	Herramienta que detecta hechos noticiables dentro de un texto.	Implementado
P3	Claim hunter	Herramienta que detecta hechos noticiables dentro de un tuit de la red social Twitter.	Implementado
P4	Claim similarity	Herramienta que detecta si una información ha sido verificada por otro <i>fact checker</i> en cualquier otro idioma.	Implementado
P5	Detección de clickbait	Detecta titulares que puedan considerarse <i>clickbait</i> . Una métrica que formará parte de un proyecto más ambicioso.	En desarrollo
P6	Detección de Toxicidad política	Mide, según la información contenida en un tuit, su nivel de toxicidad política.	Paralizado

Fuente: Elaboración propia.

P1. Transcripción de audio a texto

Este servicio fue una de las primeras soluciones de IA en ser implementada en la redacción de Newtral, aproximadamente en 2018. A diferencia del resto de proyectos, este no es de creación propia, aunque sí ha tenido una serie de modificaciones. La transcripción de audio a texto es una de las herramientas más útiles para los periodistas, ya que les permite analizar de forma mucho más rápida un discurso. Además, integra al P2, lo que facilita la detección de los hechos verificables mientras se lee el texto.

E2 describe su uso como una ayuda para la redacción, no sólo para reducir el tiempo que implica una transcripción hecha a mano, sino también para detectar fácilmente los hechos verificables.

La IA de la transcripción procede de un tercero, la empresa británica Speechmatics, puntera en software de reconocimiento del discurso. El programa se ha integrado en el CMS documental de Newtral como la primera parte del *pipeline* (E1). A su vez, el transcriptor se ha adaptado para introducirse en un editor creado por los ingenieros de Newtral, que es donde se integra la IA de Claim Detection (P2).

P2. Claim Detection

El segundo proyecto es un detector de hechos verificables (Claim Detection). Está estrechamente relacionado con el P1, pues ambos forman parte de la API de edición. Además, el P2 es esencial en el funcionamiento y el desarrollo del resto de proyectos.

La IA del Claim Detection subraya automáticamente frases que puedan ser verificables. Este proceso lo hace de forma multilingüe y puede reconocer casi cualquier idioma pese a estar entrenada solo con frases en castellano. El proceso de entrenamiento comenzó con el reconocimiento de 10.000 frases; hoy en día, la IA llega a las 500.000. Las pruebas para comprobar su correcto funcionamiento se realizaron con 21 idiomas de la UE. Detecta idiomas extranjeros con un 70-75% de precisión, mientras que la fidelidad en castellano oscila entre el 85 y el 90%.

El uso conjunto del transcriptor y el Claim Detector se ha adaptado totalmente al flujo de trabajo de los periodistas verificadores, que monitorizan a diario las principales intervenciones de actores políticos en medios de comunicación con estas herramientas. A medida que reproducen el audio o el video de una intervención, el algoritmo de Claim Detection señala en la transcripción las frases que puedan ser verificables.

Los periodistas controlan así doblemente el discurso evitando dejar algo sin revisar: se repasa el texto escrito, más frío y fácil de analizar, y, a su vez, se controla el lenguaje no verbal del emisor y el contexto en el que se encuentra (E2). Los fallos detectados en la transcripción se corrigen para así reentrenar a la IA.

P3. Claim Hunter

El tercer proyecto, Claim Hunter, busca en varios centenares de cuentas de Twitter (en el momento de la investigación aún no era X) frases que puedan ser verificables.

Desde el punto de vista computacional, el Claim Hunter es lo mismo que el Claim Detection; la diferencia está en el trato que el usuario le da, ya que cada una de estas soluciones de IA posee una interfaz y un modo de uso distinto.

El equipo de verificación del discurso político destaca que este software monitoriza alrededor de 200 y 300 cuentas. Los usuarios son desde diputados hasta representantes de partidos políticos y presidentes y consejeros autonómicos.

La IA detecta lo verificable y lo envía a un canal de Slack. Una vez ahí, el equipo lo revisa y decide si se descarta o si, por el contrario, se investiga para desmentirlo o matizarlo. De media, la IA informa de unos 70 u 80 tuits al día contra los 500 de media que podrían tener que revisarse en caso de no tener el proceso automatizado. La IA sirve, como destacan los periodistas, para detectar argumentarios de partidos y saber qué políticos repiten día a día informaciones ya desmentidas o carentes de una explicación mayor y necesaria.

P4. Claim Similarity

El P4 es un buscador que comprueba si una afirmación ha sido verificada ya. Estos hechos verificados se encuentran en una API de fact-checkers que proporciona Google, conocida como Fact Check Explorer (E4).

P4 funciona en más de 150 idiomas distintos, aunque aún no se haya perfeccionado más que con unos cuantos (“español e hindi sobre todo” (E1)). Este algoritmo usa complejas técnicas computacionales que le hacen a ser plurilingüe (E4). En el futuro se espera que este proyecto pueda salir para el uso del público general y no sólo para uso interno.

P5. Detección de clickbait

El proyecto de IA más reciente de Newtral es el de detección de titulares con *clickbait*. Estos tienden a repetir ciertos patrones y expresiones que hacen posible que una IA pueda detectar y rastrear para denunciarlos. En el momento de la investigación, el proyecto se encontraba en su fase de desarrollo.

Este P5 se ha ido desarrollando casi de forma exclusiva por una trabajadora que entró en Newtral en septiembre de 2021 (E3). El proyecto cuenta con 200.000 unidades de datos y emplea IA de última generación (transformer) para agilizar procesos. Este proyecto formará parte de uno más grande que trate de detectar si la información es

confiable o no. El *clickbait* es solo una primera aproximación a ello, pero se espera que de cara al futuro se vayan ampliando las métricas que se aplican, como por ejemplo si hay polarización o si prima la subjetividad sobre la objetividad.

P6. (Interrumpido) Detección de toxicidad política

El proyecto de detección de la toxicidad política se encuentra, a fecha de la investigación, interrumpido indefinidamente ya que no podía asegurarse su objetividad (E1). Durante la etapa de I+D+i, se creó como modelo, pero no llegó a diseñarse como producto.

La finalidad de este proyecto es detectar, en tuits de políticos, índices cuantificables de toxicidad. Estos índices varían dependiendo de si los mensajes incluyen discurso del odio, si buscan crispar o si se hablaba en términos de confrontación entre grupos políticamente opuestos. El resultado debía ser una puntuación que valorase como buena o mala una información.

Otros proyectos

Junto a los 6 proyectos que se han considerado principales, el equipo de Newtral emplea IA mucho más simple y sencilla para otros proyectos puntuales relacionados con investigaciones y formatos innovadores. Dentro de esta categoría destacan el reportaje “Las calles de ellas” (Boira et al., 2022) o un proyecto de newsgame, en curso durante la investigación. El uso de la IA en “Las calles de ellas” fue para clasificar nombres de hombre y de mujer de distintas calles de España en base a asociaciones automáticas. En el newsgame, por otro lado, la IA es mucho más sencilla y no implica ningún tipo de Machine Learning, sino que se programan “una serie de reglas que determinan de forma heurística el comportamiento de la máquina” (E1).

5. DISCUSIÓN Y CONCLUSIONES

Los resultados de este artículo profundizan en cómo la aplicación de la IA en un medio especializado en verificación como Newtral ha dado pie a la creación de proyectos innovadores y equipos específicamente dedicados al desarrollo de soluciones

tecnológicas muy cercanas a los periodistas. Se han identificado cuatro proyectos de IA en Newtral (RQ1): una herramienta de transcripción automática de textos (P1), un detector de hechos noticiables dentro de un texto (P2), un programa similar al anterior, pero para detectar textos noticiables en Twitter (P3) y una herramienta para comprobar si un bulo ha sido verificado con anterioridad en cualquier idioma (P4). A estos proyectos hay que añadir un detector de titulares con elementos de clickbait (P5), en desarrollo en el momento de la investigación, y un detector de toxicidad política en mensajes publicados en Twitter (P6), paralizado por los problemas éticos que generaba al no poder asegurar su objetividad. Junto a estos usos, la IA se aplica para generar productos periodísticos especiales, ya sea en lo genuino de la metodología de investigación (reportaje sobre las calles con nombre de mujer) o en su formato de visualización (newsgame).

Estos proyectos comparten características comunes con los de otros medios españoles e internacionales que han aplicado la IA, como los proyectos 2, 3 y 4 y sus similitudes con FactStream del Duke Reporter's Lab (Beckett & Yaseen, 2023) entre otros. No obstante, en el estudio no se ha observado ningún proyecto que busque hechos verificables en toda una red social como el ejemplo del News Tracer de Reuters, aunque el proyecto 3 sea similar (con un número limitado de perfiles).

La IA desarrollada en Newtral tiene aplicaciones diversas que ayudan a los periodistas (RQ2) en tres áreas principales: recopilación, producción y difusión (Beckett & Yaseen, 2023). Dentro de estas categorías, se encuentran labores ya estudiadas por varios autores en relación con otros medios (p.ej. Canavilhas, 2022; Mondría-Terol, 2023), como la detección de temas relevantes, recopilación de información y transcripción automática, mientras que queda de lado la generación automática de contenido, utilidad menos explorada por los verificadores (Sánchez-González et al., 2022). Al no hacer uso de IA para generar contenido, los periodistas mantienen su autoridad sobre todo aquello que se publique; por el otro lado, tareas como la identificación automática de hechos noticiables en un texto no pueden reemplazar el criterio humano por las

implicaciones sociopolíticas difíciles de discernir por una computadora, lo que resta autonomía a la máquina (Stray, 2019).

La interacción entre profesionales pertenecientes a los campos de la ingeniería y el periodismo (RQ3) funciona de manera eficaz, gracias a la eficiente comunicación, la habilidad para adaptarse a las distintas disciplinas y las reuniones conjuntas. Esta sinergia laboral, que fluye sin obstáculos, influye positivamente en la génesis y el desarrollo de nuevos proyectos, que ejemplifican la “colaboración interdisciplinaria” y la “cultura de aprendizaje continuo” (Sánchez-González et al., 2022). Entronca además en el contexto del desarrollo de proyectos relacionados con la IA y la mutación de la cultura profesional de técnicos y periodistas hacia una visión que facilite la difusión de las innovaciones en la organización (García-Avilés et al., 2018).

La implementación de la IA en Newtral se da en un contexto posterior a la pandemia, en el que los medios han acelerado su transformación (García-Avilés et al. 2022), apostando por una tecnología capaz de “elevar la eficiencia productiva” (Carvajal et al, 2022, p.13) y creando una relación complementaria y una integración de la IA en el seno de la redacción. Esto se verá potenciado por la cercanía entre los equipos y, como ha sucedido con todas las grandes innovaciones, acabará mutando todos los procesos de los profesionales (van Dalen, 2012).

La ética está muy presente en la relación de los periodistas con la IA (RQ4). Los profesionales entrevistados son conscientes de que, especialmente en un verificador homologado internacionalmente, deben respetarse férreos principios deontológicos. Los sesgos de los algoritmos pueden solventarse aumentando el entrenamiento de la máquina (con el coste que ello conlleva), pero que por el momento seguirán existiendo. El uso de técnicas *human-in-the-loop* garantiza que cualquier contenido manipulado con IA haya pasado antes por la revisión de un profesional responsable antes de su publicación. Esta técnica dibuja lo que Johnson (2023, p. 1) llama “una línea en la arena” para los periodistas: al impedir la publicación automatizada, los humanos tienen la última palabra en cuanto a criterios de publicación y la autoridad para rendir cuentas por él. Sin embargo, junto a la automatización de los procesos vienen aparejadas una

serie de problemas éticos todavía sin solucionar, como la dependencia de fondos y recursos provenientes de grandes empresas tecnológicas, que aumentan la subordinación de los medios de comunicación a estas corporaciones, que aumentan su dominio en la industria (Simon, 2022).

La investigación tiene una serie de limitaciones. El trabajo se basa en un único caso que no puede extrapolarse a la industria mediática española ni tampoco a todos los medios que han incorporado la IA. Además, la muestra es bastante reducida: tan solo se han realizado cuatro entrevistas y sólo una de ellas a un periodista. Finalmente, no se ha hecho un estudio longitudinal para mostrar la evolución a través de un periodo de tiempo, sino que la investigación es una suerte de foto fija.

6. REFERENCIAS

Aguado-Guadalupe, G., & Bernaola-Serrano, I. (2020). Verificación en la infodemia de la Covid-19. El caso Newtral. *Revista Latina de Comunicación Social*, 78, 289-308.

<https://doi.org/10.4185/RLCS-2020-1478>

Almansa-Martínez, A., Fernández-Torres, M. J., & Rodríguez-Fernández, L. (2022). Desinformación en España un año después de la COVID-19. Análisis de las verificaciones de Newtral y Maldita. *Revista Latina de Comunicación Social*, 80, 183-200. <https://doi.org/10.4185/RLCS-2022-1538>

Ananny, M. (2016). Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science Technology and Human Values*, 41(1), 93-117.

<https://doi.org/10.1177/0162243915606523>

Beckett, C., & Yaseen, M. (2023). *Generating Change. A global survey of what news organisations are doing with AI*. <https://cutt.ly/ow0ZgJ4I>

Beltrán, J., Míguez, R., & Larraz, I. (2021). *ClaimHunter: An unattended tool for automated claim detection on Twitter*. KnOD'21 Workshop. <https://cutt.ly/6eskkZjQ>

- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication, 33*(2), 122-139. <https://doi.org/10.1177/0267323118760317>
- Boira, P., Pita, C., & Larraz, I. (2022, enero 25). *Las calles de ellas*. Newtral. <https://cutt.ly/Jw0Zg5mX>
- Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C. H. (2019). Artificial Intelligence and Journalism. *Journalism and Mass Communication Quarterly, 96*(3), 673-695. <https://doi.org/10.1177/1077699019859901>
- Calvo-Rubio, L. M., & Ufarte-Ruiz, M. J. (2021). Artificial intelligence and journalism: Systematic review of scientific production in web of science and scopus (2008-2019). *Communication and Society, 34*(2), 159-176. <https://doi.org/10.15581/003.34.2.159-176>
- Cameron, K. S., & Quinn, R. E. (2011). *Diagnosing and Changing Organizational Culture: Based on the Competing Values Framework* (3rd ed.). Wiley.
- Canavilhas, J. (2022). Artificial intelligence in journalism: Automatic translation and recommendation system in the project “A European Perspective” (EBU). *Revista Latina de Comunicacion Social, 80*, 1-13. <https://doi.org/10.4185/RLCS-2022-1534>
- Carvajal, M., Mondéjar, D., Valero-Pastor, J. M., De-Lara, A., García-Avilés, J. A., & Arias-Robles, F. (2022). The most prominent journalistic innovations in Spain (2010-2020): organizational, industrial and social characteristics and impact. *Profesional de la Informacion, 31*(3). <https://doi.org/10.3145/epi.2022.may.04>
- Cazalens, S., Leblay, J., Lamarre, P., Manolescu, I., & Tannier, X. (2018). Computational fact checking. *Proceedings of the VLDB Endowment, 11*(12), 2110-2113. <https://doi.org/10.14778/3229863.3229880>
- Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data and Society, 4*(2). <https://doi.org/10.1177/2053951717718855>

Cobisa, I. (2022). *Inteligencia Artificial como palanca de generación de valor. Madurez del mercado en España*. <https://cutt.ly/Ow0Zhlpn>

de Oliveira, N. R., Pisa, P. S., Lopez, M. A., de Medeiros, D. S. V., & Mattos, D. M. F. (2021). Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. *Information*, 12(1), 38. <https://doi.org/10.3390/info12010038>

Demartini, G., Mizzaro, S., & Spina, D. (2020). Human-in-the-loop Artificial Intelligence for Fighting Online Misinformation: Challenges and Opportunities. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 43(3), 65-74. <https://cutt.ly/Besej4MD>

Diakopoulos, N. (2015). Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398-415. <https://doi.org/10.1080/21670811.2014.976411>

Dörr, K. N., & Hollnbuchner, K. (2017). Ethical Challenges of Algorithmic Journalism. *Digital Journalism*, 5(4), 404-419. <https://doi.org/10.1080/21670811.2016.1167612>

Fernandes, E., Moro, S., & Cortez, P. (2023). Data Science, Machine learning and big data in Digital Journalism: A survey of state-of-the-art, challenges and opportunities. *Expert Systems with Applications*, 221, 119795. <https://doi.org/10.1016/j.eswa.2023.119795>

García-Vivero, G., & López, X. (2021). La verificación de datos en Europa. Análisis de 5 iniciativas europeas: Maldita.es, Newtral, Pagella Política, Les Décodeurs y BBC Reality Check. *adComunica*, 21, 235-264. <https://doi.org/10.6035/2174-0992.2021.21.12>

García-Avilés, J. A., Arias-Robles, F., de Lara-González, A., Carvajal, M., Valero-Pastor, J. M., & Mondéjar, D. (2022). How COVID-19 is Revamping Journalism: Newsroom Practices and Innovations in a Crisis Context. *Journalism Practice*. <https://doi.org/10.1080/17512786.2022.2139744>

García-Avilés, J. A., Carvajal-Prieto, M., Arias, F., & De Lara-González, A. (2018). How journalists innovate in the newsroom. Proposing a model of the diffusion of innovations in media outlets. *The Journal of Media Innovations*, 5(1), 1-16.

<https://doi.org/10.5617/jomi.v5i1.3968>

García-Marín, D. (2022). Modelos algorítmicos y fact-checking automatizado. Revisión sistemática de la literatura. *Documentación de las Ciencias de la Información*, 45(1), 7-16. <https://doi.org/10.5209/dcin.77472>

García-Orosa, B., Canavilhas, J., & Vázquez-Herrero, J. (2023). Algorithms and communication: A systematized literature review. *Comunicar*, 30(74), 9-21.

<https://doi.org/10.3916/C74-2023-01>

Goodman, L. A. (1961). Snowball Sampling. *The Annals of Mathematical Statistics*, 32(1), 148-170. <https://doi.org/10.1214/aoms/1177705148>

Graves, L. (2018). *Understanding the Promise and Limits of Automated Fact-Checking*.

<https://cutt.ly/Hw0ZlQRV>

Haim, M., & Graefe, A. (2017). Automated News. *Digital Journalism*, 5(8), 1044-1059.

<https://doi.org/10.1080/21670811.2017.1345643>

Hassan, N., Adair, B., Hamilton, J. T., Li, C., Tremayne, M., Yang, J., & Yu, C. (2015). *The Quest to Automate Fact-Checking*. <https://cutt.ly/3w0ZlJVj>

Johnson, P. R. (2023). A Case of Claims and Facts: Automated Fact-Checking the Future of Journalism's Authority. *Digital Journalism*.

<https://doi.org/10.1080/21670811.2023.2174564>

Kim, D., & 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>

Larraz, I., Míguez, R., & Sallicati, F. (2023). Semantic similarity models for automated fact-checking: ClaimCheck as a claim matching tool. *Profesional de la información*, 32(3), e320321. <https://doi.org/10.3145/epi.2023.may.21>

Lewis, S. C., Sanders, A. K., & Carmody, C. (2019). Libel by Algorithm? Automated Journalism and the Threat of Legal Liability. *Journalism & Mass Communication Quarterly*, 96(1), 60-81. <https://doi.org/10.1177/1077699018755983>

Lewis, S. C., & Zamith, R. (2017). On the Worlds of Journalism. En *Remaking the News* (pp. 111-128). The MIT Press. <https://doi.org/10.7551/mitpress/10648.003.0012>

López-García, X., & Vizoso, Á. (2021). High-tech journalism: A sign of the digital era of the third millennium. *Profesional de la Información*, 30(3), e300301. <https://doi.org/10.3145/epi.2021.may.01>

López-Marcos, C., & Vicente-Fernández, P. (2021). Fact Checkers Facing Fake News and Disinformation in the Digital Age: A Comparative Analysis between Spain and United Kingdom. *Publications*, 9(3), 36. <https://doi.org/10.3390/publications9030036>

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. *El Profesional de la información*, 32(5), e320520. <https://doi.org/10.3145/epi.2023.sep.20>

Maldita.es lanza el primer chatbot de Whatsapp en español para verificar bulos automáticamente. (2020, junio 17). Maldita.es. <https://cutt.ly/gw0Zl9fU>

Maslej, N., Fattorini, L., Kiesow Cortez, E., Ngo, H., & Rahman, R. (2023). *Artificial Intelligence Index Report 2023*. <https://cutt.ly/hw0ZzaMN>

Meléndez-Malavé, N., & Pérez-Quintana, P. (2023). La verificación de bulos relacionados con el feminismo: Un análisis de los desmentidos sobre género publicados en Maldita.es y Newtral. *Signo y Pensamiento*, 42. <https://doi.org/10.11144/Javeriana.syp42.vbrf>

Mondría-Terol, T. (2023). Innovación Mediática: aplicaciones de la inteligencia artificial en el periodismo en España. *Textual & Visual Media*, 17(1), 41-60.

<https://doi.org/10.56418/txt.17.1.2023.3>

Oremus, W. (2014, marzo 17). *The First News Report on the L.A. Earthquake Was Written by a Robot*. Slate. <https://cutt.ly/dwOZzn7i>

Parratt-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*, 30(3). <https://doi.org/10.3145/epi.2021.may.17>

Pérez-Escoda, A., Pedrero-Esteban, L. M., Rubio-Romero, J., & Jiménez-Narros, C. (2021). Fake news reaching young people on social networks: Distrust challenging media literacy. *Publications*, 9(2). <https://doi.org/10.3390/publications9020024>

Porcu, O., Hermans, L., & Broersma, M. (2020). Unlocking the Newsroom: Measuring Journalists' Perceptions of Innovative Learning Culture. *Journalism Studies*, 21(10), 1420-1438. <https://doi.org/10.1080/1461670X.2020.1758956>

Rao, A., & Verweij, G. (2022). *Sizing the prize. PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution*. <https://cutt.ly/dwOZZIRn>

Rojas-Caja, F. (2020). *El fact checking. Las agencias de verificación de noticias en España*. <https://cutt.ly/0wOZP7JS>

Rojas-Torrijos, J. L., & Toural-Bran, C. (2019). Automated sports journalism. The AnaFut case study, the bot developed by El Confidencial for writing football match reports. *Doxa Comunicacion*, 2019(29), 235-254. <https://doi.org/10.31921/doxacom.n29a12>

Sánchez-González, M., Sánchez-Gonzales, H. M., & Martínez-Gonzalo, S. (2022). Inteligencia artificial en verificadores hispanos de la red IFCN: proyectos innovadores y percepción de expertos y profesionales. *Estudios sobre el Mensaje Periodístico*, 28(4), 867-879. <https://doi.org/10.5209/esmp.82735>

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>

Stray, J. (2019). Making Artificial Intelligence Work for Investigative Journalism. *Digital Journalism*, 7(8), 1076-1097. <https://doi.org/10.1080/21670811.2019.1630289>

Thurman, N., Dörr, K., & Kunert, J. (2017). When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, 5(10), 1240-1259. <https://doi.org/10.1080/21670811.2017.1289819>

Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, Automation, and News. *Digital Journalism*, 7(8), 980-992. <https://doi.org/10.1080/21670811.2019.1685395>

Túñez-López, J. M., & Tejedor-Calvo, S. (2019). Artificial intelligence (AI) and journalism [monograph presentation]. *Doxa Comunicacion*, 2019(29), 163-168.

<https://doi.org/10.31921/doxacom.n29a8>

Túñez-López, J.-M., Fieiras-Ceide, C., & 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>

Ufarte-Ruiz, M. J., & Manfredi-Sánchez, J. L. (2019). Algorithms and bots applied to journalism. The case of Narrativa Inteligencia Artificial: structure, production and informative quality. *Doxa Comunicación. Revista interdisciplinar de estudios de comunicación y ciencias sociales*, 29, 213-233.

<https://doi.org/10.31921/doxacom.n29a11>

Valles, M. S. (2002). *Entrevistas cualitativas*. Centro de Investigaciones Sociológicas.

van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, 6(5-6), 648-658.

<https://doi.org/10.1080/17512786.2012.667268>

Vázquez-Herrero, J., Vizoso, Á., & López-García, X. (2019). Innovación tecnológica y comunicativa para combatir la desinformación: 135 experiencias para un cambio de rumbo. *El Profesional de la Información*, 28(3).

<https://doi.org/10.3145/epi.2019.may.01>

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

Waisbord, S. (2018). Truth is What Happens to News: On journalism, fake news, and post-truth. *Journalism Studies*, 19(13), 1866-1878.

<https://doi.org/10.1080/1461670X.2018.1492881>

Wu, S., Tandoc, E. C., & Salmon, C. T. (2019). Journalism Reconfigured. *Journalism Studies*, 20(10), 1440-1457. <https://doi.org/10.1080/1461670X.2018.1521299>

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2021). A Survey of Human-in-the-loop for Machine Learning. <https://doi.org/10.1016/j.future.2022.05.014>

Zarocostas, J. (2020). How to fight an infodemic. *The Lancet*, 395(10225), 676.

[https://doi.org/10.1016/S0140-6736\(20\)30461-X](https://doi.org/10.1016/S0140-6736(20)30461-X)

Zheng, Y., Zhong, B., & Yang, F. (2018). When algorithms meet journalism: The user perception to automated news in a cross-cultural context. *Computers in Human Behavior*, 86, 266-275. <https://doi.org/10.1016/j.chb.2018.04.046>