

INTELIGENCIA ARTIFICIAL GENERATIVA Y CONFIANZA EN LOS MEDIOS. UN ANÁLISIS DE LA DETECCIÓN DE IA EN NOTICIAS USANDO GPTZERO

Generative artificial intelligence and media trust. An
analysis of AI detection in news using GPTZero

 **Francisco Leslie López del Castillo Wilderbeek**¹: Universitat Pompeu Fabra,
España.

franciscoleslie@alumni.upf.edu

Cómo citar el artículo:

López del Castillo Wilderbeek, Francisco Leslie. (2025). Inteligencia artificial generativa y confianza en los medios. Un análisis de la detección de IA en noticias usando GPTzero [Generative artificial intelligence and media trust. An analysis of AI detection in news using GPTZero]. *Vivat Academia. Revista de Comunicación*, 158, 1-17. <https://doi.org/10.15178/va.2025.158.e1556>

RESUMEN

Introducción: El uso de la inteligencia artificial (IA) en la generación de noticias es una realidad que puede poner en peligro la confianza en los medios de comunicación.

Metodología: El presente trabajo ha tenido el objetivo de indagar mediante análisis de contenido si la mejor plataforma para la detección de uso de IA en textos: GPTZero, es garantía para discriminar entre noticias elaboradas completamente con IA y contenidos firmados por profesionales. Para ello se analizaron textos de noticias reales (prensa y digital) y de noticias elaboradas por inteligencia artificial (ChatGPT).

Resultados: Los resultados obtenidos han indicado que GPTZero ofrece una baja fiabilidad para detectar textos generados con ChatGPT (7,3%) y, que puede señalar como falsos positivos contenidos que no han sido elaborados con inteligencia artificial.

Discusión: Los análisis realizados con dos algoritmos diferentes indicaron que

¹ **Francisco Leslie López del Castillo Wilderbeek:** Doctor en comunicación por la Universitat Pompeu Fabra, Profesor colaborador en la UOC, Profesor asociado en la UOC. Ejerce como analista de medios de comunicación en la empresa Rebold.

GPTZero no descubrió el uso completo de ChatGPT en un corpus de noticias y, de la misma forma, advirtió de cierta probabilidad de uso colaborativo de IA (redacción por un humano con soporte de IA) en noticias que fueron firmadas por periodistas profesionales. **Conclusiones:** Los resultados obtenidos han generado más dudas que certezas sobre si GPTZero puede ayudar a señalar el uso de IA o asegurar el prestigio de los medios de comunicación.

Palabras clave: medios de comunicación, inteligencia artificial, GPTZero, ChatGPT, prensa, cibermedios.

ABSTRACT

Introduction: The use of AI in news generation is a reality that can jeopardize trust in the media. **Methodology:** The aim of this paper was to analyze through content analysis whether the best platform for detecting the use of AI in texts: GPTZero, is guaranteed to discriminate between news completely elaborated with AI and contents signed by professionals. For this purpose, texts of real news (press and digital) and news elaborated by artificial intelligence (ChatGPT) were analyzed. **Results:** The results obtained indicated that GPTZero offers a low reliability for detecting texts generated with ChatGPT (7.3%) and that it can signal as false positives contents that have not been elaborated with artificial intelligence. **Discussion:** The analyses performed with two different algorithms indicated that GPTZero did not discover the full use of ChatGPT in a corpus of news and, likewise, warned of a certain probability of collaborative use of AI (writing by a human with AI support) in news that were signed by professional journalists. **Conclusions:** The results obtained have generated more doubts than certainties about whether GPTZero can help signal the use of AI or ensure the prestige of the media.

Keywords: media, artificial intelligence, GPTZero, ChatGPT, press, cyber media.

1. INTRODUCCIÓN

1.1. Inteligencia Artificial y producción de noticias

Tal como advierte el Reuters Institute los medios de comunicación se encuentran en un proceso de integración de la Inteligencia Artificial en sus productos. Concretamente el 28% de medios encuestados afirma que ya trabajan con este tipo de tecnología mientras que un 39% están haciendo pruebas para su incorporación (Newman, 2023).

Dentro del concepto paraguas de la IA una variante muy relevante es la IA generativa puesto que esta tecnología es capaz de elaborar contenidos como texto, imágenes o incluso código de programación tomando como referente conocimiento previo y gracias a unas sucintas instrucciones. De esta forma, la generación de contenido plausiblemente novedoso supone un salto muy significativo ya que puede confundirse con material creado por el ser humano y que sí que es material original.

Aplicar esta tecnología en la producción de noticias es una realidad para los medios de comunicación ya que conlleva un ahorro de costes (Aramburú Moncada *et al.*, 2023)

pese a la necesidad de contratar a personal especializado (de-Lima-Santos y Salaverría, 2021). Sin embargo, también fomenta un problema añadido ya que las noticias son un bien público de forma que pueden poner en alerta a los organismos reguladores (Sandrini y Somogyi, 2023). De la misma manera, y obviando la presumible neutralidad inherente a la función de los medios de comunicación, el perfeccionamiento de estos procesos puede conducir a la proliferación de *deepfakes* (imágenes falsas de alta credibilidad) y, muy especialmente, la difusión de desinformación (Shoaib *et al.*, 2023). Esto es debido a que la IA generativa, entrenada con grandes modelos textuales, puede crear noticias falsas sin ninguna dificultad, que sean coherentes (Botha y Pieterse, 2020) y con un estilo casi idéntico al de los auténticos profesionales pese a que el resultado final no sea necesariamente veraz (Franganillo, 2023). Según una encuesta realizada por IPSOS en 29 países el 74% de la población cree que la IA hace más fácil la generación de *fake news* creíbles (Dunne, 2023).

Además de esto, el trabajo de Longoni *et al.* (2022) llegó a la conclusión de que el uso reconocido de la IA por parte de los productores de noticias puede ser perjudicial pese a que el contenido sea cierto. Según estos autores si los medios divulgan que usan IA para sus contenidos están dañando la confianza y credibilidad que los receptores tienen en los productores de noticias, aunque estas sean veraces. Concretamente llegaron a afirmar que “una consecuencia importante de nuestros experimentos es que los llamamientos a la transparencia en el uso de la IA pueden ser contraproducentes” (Longoni *et al.*, 2022, p. 102).

Por tanto, el uso de la IA por parte de las productoras de noticias genera un doble peligro. Por una parte, la IA generativa se encuentra fuertemente relacionada con la elaboración masiva de noticias falsas. Por otra, la receptividad de las audiencias no es la misma si estas son conscientes de que están ante productos que han sido elaborados, total o parcialmente, con la ayuda de IA.

1.2. Detección de noticias generadas con IA

El uso de la IA para la producción de noticias es controvertido en sí mismo. Para paliar el impacto de las noticias falsas generadas por IA existen varias propuestas que se encuentran en sintonía con los procesos ya usados en la detección de noticias falsas elaboradas manualmente.

Según la opinión de Shoaib *et al.* (2023) es de vital importancia la necesidad de desarrollar soluciones tecnológicas más sofisticadas, la formulación de políticas internacionales flexibles al entorno tecnológico y el impulso de la alfabetización mediática. Con relación al volumen de la producción de noticias falsas por IA, Deepak (2023) afirma que es necesario un perfeccionamiento de la automatización en la detección de noticias falsas al mismo nivel que la IA generativa mejora sus procesos productivos. El uso de aplicaciones para la detección del uso de la IA permitiría trascender o complementar la función de verificación humana (*fact-checking*) y, en algunos casos, evitaría sesgos subjetivos (Rodríguez-Pérez *et al.*, 2023) en los supuestos en los que esta tecnología se haya empleado para engañar a los receptores. Aunque conviene reseñar que no se trataría de una verificación de hechos en sentido estricto

ya que este proceso se encuentra relacionado con la “capacidad de detectar afirmaciones, recuperar pruebas pertinentes, evaluar la veracidad de cada afirmación y ofrecer justificación de las conclusiones aportadas” (DeVerna *et al.*, 2023, p. 2). Lo que la tecnología IA puede ofrecer es desvelar si esta ha sido usada, ya sea de forma legítima o ilegítima pero no reemplazar la cognición, el juicio y la inteligencia emocional humana.

Sin embargo, más allá de la detección del uso de IA en noticias que realmente han usado esta tecnología como soporte, subyace la cuestión sobre si un contenido elaborado con IA por un productor de prestigio (por ejemplo, un medio de comunicación que produce además *offline*) se ve afectado negativamente por la percepción que el público tiene sobre el uso de la IA en las noticias que consume.

Por este motivo las aplicaciones o programas dedicados al desvelamiento del uso de IA para la generación de noticias podrían lesionar la reputación de los medios de comunicación que usen esta tecnología que como se ha visto ya está siendo incorporada en los grupos editoriales. Es más, en los casos en los que se detectara incorrectamente el uso de IA (falso positivo) los medios afectados podrían ser acusados falsamente y, por tanto, presumiblemente perder la confianza de su audiencia tal como afirma el trabajo de Longoni *et al.* (2022).

No es casual que según los resultados del I Estudio sobre la desinformación en España (Sádaba-Chalezquer y Salaverría-Aliaga, 2022) el 84,6% de los encuestados mostrara su preferencia hacia los medios de comunicación tradicionales para informarse en detrimento de las redes sociales ya que los primeros poseen “equipos profesionales de periodistas que verifican, contrastan y analizan las informaciones” (Sádaba-Chalezquer y Salaverría-Aliaga, 2022, p. 11). La cuestión radicaría en que esta confianza se vería fuertemente dañada si la participación de profesionales cualificados se ve minimizada o directamente desaparece por el advenimiento de la IA generativa.

Actualmente, la referencia dentro de las aplicaciones disponibles para la detección del uso de IA en corpus textuales es el programa GPTZero. Tal como indica en su web este programa es el estándar de oro en detección de IA, entrenado para detectar ChatGPT, GPT4, Bard, LLaMa y otros modelos de IA (GPTZero, s.f.).

2. OBJETIVOS

El objetivo de la presente investigación es dilucidar si GPTZero es realmente una garantía para la detección de *fake news* generadas por IA (noticias elaboradas completamente sin basarse en un acontecimiento real), el uso de IA como apoyo por los productores de noticias y/o si supone una amenaza para los medios de comunicación tradicionales al apreciar indicios de uso de IA que puedan poner en riesgo su credibilidad como medio de referencia.

Concretamente se pretende analizar la tasa de acierto de GPTZero en corpus textuales. Por un lado, noticias reales extraídas de medios de comunicación de referencia (tanto prensa en papel como digital) y por otro, contenidos generados por la inteligencia artificial ChatGPT.

3. METODOLOGÍA

Para cumplir con el objetivo indicado anteriormente esta investigación ha analizado mediante GPTZero un corpus de noticias. Este corpus se ha dividido en dos grupos. Por un lado, noticias generadas íntegramente por la IA Chat-PGT 3.5 mediante la instrucción “Elabora una noticia sobre (NOMBRE DE ORGANIZACIÓN)”. La elección de esta instrucción fue debido a que ChatGPT no responde a la instrucción “Elabora una *fake news* sobre (NOMBRE DE ORGANIZACIÓN)”. De igual forma, estaríamos ante noticias inventadas (no basadas en acontecimientos reales) generadas por IA (*fake news*). Para confirmar este extremo este corpus de noticias se analizó manualmente y se confirmó que los hechos narrados por ChatGPT no tienen vinculación con acontecimientos reales presentes o pasados. Concretamente las organizaciones de las que se generó una noticia falsa mediante ChatGPT fueron Iberdrola, Naturgy y Acciona.

Por otro lado, se recuperó un corpus de noticias publicadas tanto en prensa en papel como en prensa digital sobre las tres organizaciones citadas anteriormente (una noticia por cada organización y por cada medio). Los medios que se analizaron en prensa tradicional fueron *Expansión*, *Cinco Días* y *El Economista* ya que son los medios económicos de referencia en España. En el caso de medios digitales se escogieron únicamente medios digitales nativos (sin matriz *offline*), los medios fueron: *El Confidencial*, *OK Diario*, *El Español* y *Vozpópuli*. La justificación de la elección es que todos estos se encuentran entre los cinco mejores medios digitales nativos según el ranking del portal de medios de comunicación PR Noticias con fecha de octubre de 2023 (PR Noticias, s.f.).

En ambos corpus se usó como expresión de búsqueda de recuperación el nombre de las tres organizaciones ya citadas con presencia en el titular (para asegurar el protagonismo de la organización dentro de la noticia analizada).

Los contenidos de los tres corpus (IA, papel y digital) se pasaron por el analizador GPTZero en fecha de 12 de enero de 2023. Este es un dato importante ya que el algoritmo de detección de IA se actualizó con fecha del 17 de enero de 2024 de forma que se obtuvieron dos resultados diferentes replicando el mismo experimento y las mismas condiciones.

De estos tres corpus, además del análisis por GPTZero, se extrajeron mediante análisis de contenido variables en las 24 unidades de análisis para buscar correlaciones con el porcentaje de posibilidad de intervención de IA. Las categorías de análisis que se extrajeron fueron: número de palabras, presencia de la organización entidad, firma del periodista (si corresponde) y tipo de fuente (nativo digital o papel tradicional).

Tal como señala Bardin (1991) el análisis de contenido permite la obtención de indicadores, en este caso tanto cuantitativos como cualitativos. Por un lado, los indicadores cuantitativos fueron el volumen de palabras totales y el volumen la palabra clave de la entidad. Por otro, los indicadores cualitativos fueron la firma del periodista (variable dicotómica en las que se da o no se da una condición) y tipo de fuente (variable cualitativa nominal).

Tabla 1

Cuadro con la metodología aplicada en el análisis de contenido

<i>Categoría de análisis</i>	<i>Indicador</i>	<i>Tipo de variable</i>
<i>Volumen del texto</i>	<i>Cuantitativo</i>	<i>Cuantitativo</i>
<i>Volumen palabra clave entidad</i>	<i>Cuantitativo</i>	<i>Cuantitativo</i>
<i>Firma del periodista</i>	<i>Cualitativo</i>	<i>Variable dicotómica</i>
<i>Tipo de fuente</i>	<i>Cualitativo</i>	<i>Variable cualitativa nominal</i>

Fuente: Elaboración propia.

4. RESULTADOS

4.1. Primer análisis GPTZero (12 de enero de 2024)

Tal como se ha comentado se analizaron los tres corpus de contenidos usando GPTZero. Los resultados sobre implicación de IA en la generación de las noticias se expresaron en forma de porcentaje de posibilidades de que ese texto hubiera sido elaborado exclusivamente aplicando IA generativa.

Tabla 2

Resultados análisis GPTZero 12 de enero de 2024

<i>Organización</i>	<i>Chat GPT</i>	<i>Cinco Días</i>	<i>Expansión</i>	<i>El Econo mista</i>	<i>El Confide ncial</i>	<i>OK Diario</i>	<i>Vozpó puli</i>	<i>El Español</i>
<i>Iberdrola</i>	9%	5%	9%	5%	5%	5%	5%	9%
<i>Naturgy</i>	5%	5%	5%	5%	5%	5%	5%	5%
<i>Acciona</i>	8%	5%	7%	5%	7%	6%	7%	6%

Nota: Probabilidad según GPTZero (algoritmo pre 17 de enero de 2024) de que el contenido analizado haya sido elaborado íntegramente por IA. Los contenidos se segmentaron según la entidad protagonista de la noticia.

Fuente: Elaboración propia con resultados obtenidos por GPTZero (algoritmo pre 17 de enero de 2024).

Como puede verse en la Tabla 2 los resultados obtenidos con el algoritmo pre 17 de enero no distan mucho entre sí. Concretamente en el caso de noticias sobre Naturgy (tanto falsas como reales) GPTZero consideró que todas tenían las mismas posibilidades de haber sido elaboradas por IA. En los demás ejemplos los porcentajes pasaron desde la misma situación encontrada en Naturgy (mismo porcentaje de contenidos originales respecto contenido de ChatGPT) hasta valores inferiores al material creado con IA (mayor nivel de participación humana).

Resulta especialmente significativo que, mientras el corpus de IA fue generado exclusivamente con ChatGPT mediante una sencilla instrucción, la detección por parte de GPTZero no fue adecuada pues mostró un resultado máximo del 9% de posibilidades de que ese material fuera elaborado exclusivamente con IA cuando fue ese precisamente el caso. Por su parte el porcentaje mínimo fue del 5%, es decir, todavía más alejado de la realidad (aunque ChatGPT 3.5 sea una de las tecnologías que reconoce GPTZero, tal como afirma específicamente la plataforma).

Por otra parte, en el caso de los falsos positivos (detectar uso completo de IA en textos que no se han reconocido como tales por el editor) el análisis elaborado por GPTZero no fue concluyente en el sentido de que en una de las tres tipologías de noticias (Iberdrola, Naturgy y Acciona) se hallaron mismos niveles de confiabilidad sobre la participación de IA en comparación con el contenido que realmente había sido generado íntegramente por IA. Esta equiparación entre contenidos veraces y el contenido elaborado artificialmente fomenta, de nuevo, las dudas sobre la fiabilidad de la aplicación.

Tampoco se encontraron cambios relevantes entre medios de prensa en papel y medios digitales nativos, de forma que no pudo afirmarse algún tipo de preeminencia de un canal sobre otro a la hora de detectar la presencia o ausencia de tecnología de IA.

De forma más específica, la media de posibilidades sobre la generación exclusiva con IA estuvo en el 5,6% en el caso de prensa en papel y en el 5,83% en medios digitales nativos. En el caso de la proporción entre medios no se apreciaron alteraciones importantes, aunque dos medios de prensa en papel mostraron los resultados más bajos de posibilidad de generación exclusiva por IA (5% en ambos casos).

4.2. Segundo análisis GPTZero (12 de enero de 2024)

Los cambios elaborados en GPTZero con fecha del 17 de enero de 2024 se tradujeron en una alteración de los resultados obtenidos en el análisis de los textos. De forma más específica, la plataforma generó un resultado más detallado en el que no únicamente se expresaba la probabilidad de que el contenido hubiera sido generado íntegramente por IA sino también la probabilidad de que fuera un contenido elaborado en colaboración entre un humano y una inteligencia artificial. De esta forma el algoritmo aportó información de tres variables: generación enteramente por humanos (*human*), combinación de humano e IA (*mixed*) y finalmente generación enteramente por IA (*ai*), todo ello expresado en porcentajes al igual que en el algoritmo anterior.

Teniendo en cuenta estos cambios se analizaron de nuevo los tres corpus textuales.

Tabla 3

Resultados análisis GPTZero 20 de enero de 2024

Organización	ChatGPT		Cinco Días		Expansión		El Economista		El Confidencial		OK Diario		Vozpópuli		El Español	
	88	12	90	10	94	6	92	8	81	19	89	11	81	19	88	12
Iberdrola	88	12	90	10	94	6	92	8	81	19	89	11	81	19	88	12
Naturgy	84	16	90	10	87	13	89	11	88	12	90	10	90	10	89	11
Acciona	84	16	89	11	92	8	87	13	87	13	88	12	87	13	81	18

Nota: Probabilidad según GPTZero (algoritmo post 17 de enero de 2024) de que el contenido analizado haya sido elaborado íntegramente por un ser humano o en colaboración humano/IA. Los resultados referentes a generación íntegramente por IA fueron en todos los casos iguales a 0%. Los contenidos se segmentaron según la entidad protagonista de la noticia.

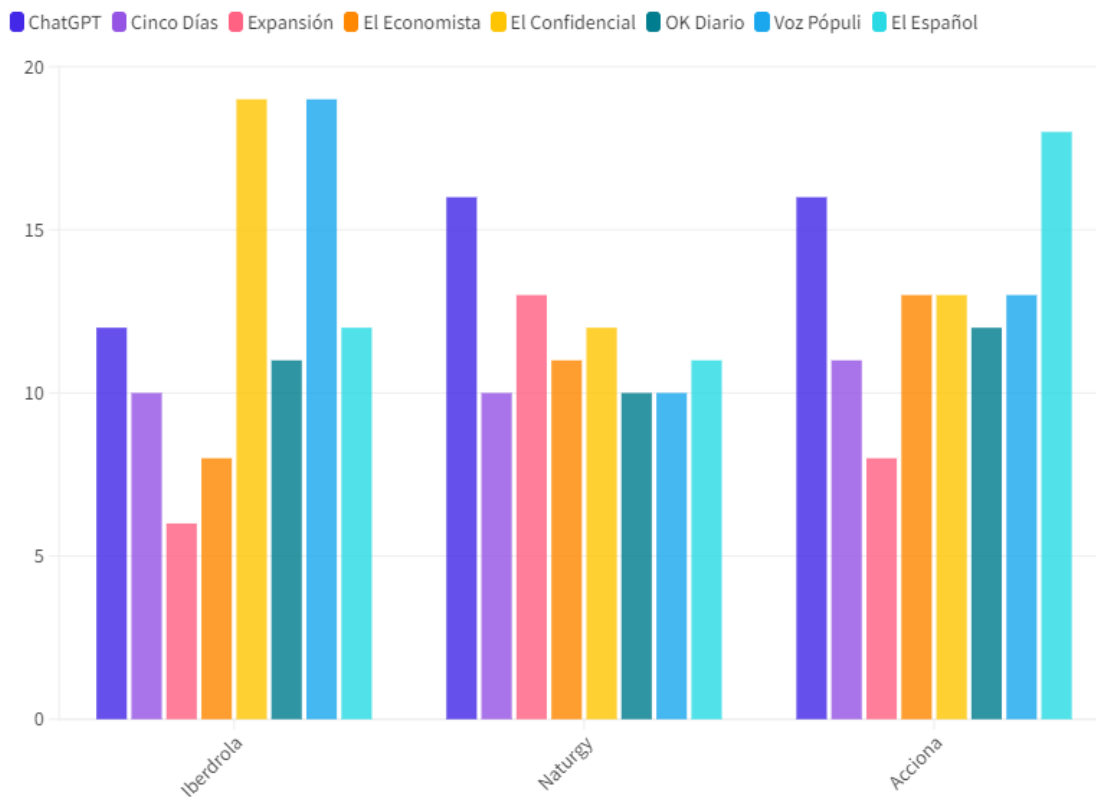
Fuente: Elaboración propia con resultados obtenidos por GPTZero (algoritmo post 17 de enero de 2024).

Como puede observarse en la Tabla 3 los resultados obtenidos fueron más detallados. Un elemento importante es que en ninguno de los contenidos se detectó que hubiera intervención exclusiva de IA (al contrario que con el algoritmo previo) y por tanto se consideró que todos los textos eran confiables pese a que tres de ellos se generaron automáticamente con ChatGPT.

A su vez, en comparación con el anterior algoritmo la tasa de acierto fue todavía inferior si tomamos la interpretación de la probabilidad de que el contenido fuera generado de forma colaborativa humano/IA. Esto es debido a que GPTZero (algoritmo post) expresó que el material elaborado con ChatGPT tenía un 12% y 16% (en dos casos) de posibilidades de ser material elaborado en colaboración. Esto, como ya se ha mencionado, es un falso negativo pues no era material colaborativo sino teóricamente original humano, pero puede servir de indicador ya que altos niveles de la variable *mixed* pueden considerarse señales de generación por IA en su totalidad. Sin embargo, los resultados sobre colaboración humano/IA fueron menores en los textos de ChatGPT que en los textos independientes tal como puede apreciarse en la Figura 1.

Figura 1

Probabilidad de que el contenido sea generado por AI según resultados análisis GPTZero 20 de enero de 2024



Fuente: Elaboración propia con datos obtenidos de GPTZero.

Concretamente se hallaron tres casos en los que GPZero detectó sólo un 81% de posibilidades de que el texto fuera íntegro humano cuando en los contenidos de ChatGPT la probabilidad más baja fue del 84%. Es decir, mientras que el algoritmo anterior encontró una equiparación errónea (niveles similares) este nuevo algoritmo llegó a dar más margen de verosimilitud a material creado por IA que a textos que presumiblemente eran trabajo de profesionales de la comunicación.

Finalmente, respecto los datos promedio, los contenidos de ChatGPT ofrecieron una media de 85,3% de posibilidades de que fuera material creado en colaboración entre humano y máquina. En las noticias de prensa en papel la media aumentó hasta el 90% y en noticias digitales de medios nativos la media fue del 86,58%. Como puede apreciarse las diferencias fueron mínimas, especialmente si se comparan las noticias digitales con las noticias generadas por ChatGPT (86,58% frente a 85,3%).

Analizando los datos de cada medio de comunicación tampoco se hallaron indicios de que un medio en concreto arrastrara o disparara los resultados generales respecto a la participación de IA en el redactado. En prensa en papel las diferencias por diario fueron imperceptibles: *Cinco Días* alcanzó un 89,6%, *Expansión* un 90% y *El Economista* un 89,3%.

Dentro de los medios digitales nativos, si bien es cierto que *OK Diario* alcanzó una ratio baja de participación de IA y cercana a los resultados de prensa en papel (89% de media para un 11% de media de probabilidad de elaboración completa por IA), el resto de los medios digitales estuvieron en rangos similares: *El Confidencial* 85,3%, *Vozpópuli* 86% y *El Español* 86%. Un detalle relevante es que las tres noticias de *OK Diario* fueron firmadas por periodistas profesionales (Carlos Ribagorda, Benjamín Santamaría y Eduardo Segovia) mientras que unas pocas del corpus de noticias digitales tenían su origen en agencias de noticias. En *El Español* la noticia sobre Iberdrola fue firmada con la expresión Agencias y en el caso de *El Confidencial* una fue firmada por EUROPA PRESS (empresa Naturgy y 12% de posibilidades de ser resultado de IA) y otra por EFE (empresa Acciona y 13% de posibilidades de ser resultado de IA). El elemento más característico sería, sin embargo, que en el subcorpus de noticias recuperadas de *El Confidencial* las noticias con menores ratios de probabilidad de generación íntegra por IA fueron aquellas en las que la autoría era de una agencia, en contraposición a la noticia con una supuesta mayor actuación de la IA (19% de posibilidades) que, en contra a la intuición general, estaba firmada por un periodista.

4.3. Análisis de contenido de noticias con menor posibilidad de ser generadas íntegramente por humanos

Para poder comprender con mayor detalle los resultados obtenidos en el punto 3.2 se aplicó el análisis de contenido a las tres noticias que fueron reconocidas con más posibilidades de haber sido generadas por una IA en colaboración con un humano (incluso más que las tres de muestra que efectivamente cumplían esta condición).

El análisis de contenido permite la comprensión de un texto de forma objetiva y sistemática cuantificando los elementos que le son característicos según unos criterios predeterminados (Neuendorf, 2017).

Las tres noticias fueron:

- 1) "Iberdrola renuncia a comprar PNM, la gran operación que EEUU vetó por el caso Villarejo"

El Confidencial / Iberdrola / 2 de enero de 2024

Posibilidad de haber sido generado por colaboración IA/humano: 81%

Posibilidad de haber sido generado íntegramente por IA: 19%

https://www.elconfidencial.com/empresas/2024-01-02/iberdrola-fusionar-avangrid-pnm-resources_3803958/

- 2) "Iberdrola eleva un 12,2% el dividendo a cuenta y amplía capital en 1.304 millones"

Vozpópuli / Iberdrola / 5 de enero de 2024

Posibilidad de haber sido generado por colaboración IA/humano: 81%

Posibilidad de haber sido generado íntegramente por IA: 19%

https://www.vozpopuli.com/economia_y_finanzas/empresas/iberdrola-eleva-dividendo-cuenta-amplia-capital-1304-millones.html

3) "Acciona anuncia que construirá una planta pionera de reciclado de palas eólicas en Navarra operativa en 2025"

El Español / Acciona / 15 de noviembre de 2023

Posibilidad de haber sido generado por colaboración IA/humano: 81%

Posibilidad de haber sido generado íntegramente por IA: 18%²

https://www.elspanol.com/invertia/empresas/energia/20231115/accion-a-anuncia-construira-planta-pionera-reciclado-palas-eolicas-navarra-operativa/809919384_0.html

La comparación entre estas tres noticias no permite, en primera instancia, extraer correlaciones sobre su pertinencia al grupo de contenidos con mayores posibilidades de haber sido generadas por IA en colaboración con un ser humano.

Por una parte, no existe uniformidad sobre las entidades protagonistas (aunque aparezcan dos noticias con entidad principal Iberdrola). Por otra, hay variedad de medios y de extensión (*El Confidencial* 814 palabras, *Vozpópuli* 474 y *El Español* 773) y de presencia de la palabra clave entidad (el término Iberdrola apareció 8 veces en *El Confidencial*, 4 veces en *Vozpópuli* y el término Acciona 6 veces en *El Español*). A su vez, existió cierto equilibrio entre la presencia de la palabra clave entidad y la extensión de cada texto (en *El Confidencial* Iberdrola apareció con un promedio de 101,75 palabras, en *Vozpópuli* con un promedio de 118,5 palabras y Acciona con un promedio de 128,8 palabras).

El condicionante más significativo podría ser que todas ellas son contenidos digitales. Teniendo en cuenta que se analizaron 12 se puede afirmar que un 25% de estas noticias fueron detectadas como potencialmente generadas por IA (en comparación con el total de corpus y teniendo en cuenta que en ningún caso GPTZero se aventuró a señalar contenidos como exclusivos de IA, si no, más bien con bajos niveles de "humanidad").

Sin embargo, el análisis de contenido también evaluó la autoría de las informaciones detectando que todas ellas estaban firmadas por profesionales del sector. Concretamente *El Confidencial* fue firmada por Miquel Roig y Javier Melguizo, *Vozpópuli* por Fernando Asunción y *El Español* por Laura Ojea.

²En este caso concreto GPTZero detectó una posibilidad del 81% de ser material elaborado por humanos, un 18% de haber sido elaborado en colaboración IA/humano y, sin embargo, el 1% restante que completaría el 100% no apareció (la posibilidad de uso completo de IA fue del 0%).

Este dato es relevante en el sentido de que estas informaciones, con presunción de mayor tasa de participación de IA, gozaban del respaldo y del prestigio no solo de un medio de comunicación sino de un grupo de profesionales (puesto que existe la posibilidad de firmar los textos con el término genérico redacción o agencia).

4.4. Análisis de contenido de noticias con mayor posibilidad de ser generadas íntegramente por humanos

Para poder comprender con mayor detalle los resultados obtenidos en el punto 3.2 se aplicó el análisis de contenido a las tres noticias que fueron reconocidas con más posibilidades de haber sido generadas exclusivamente por un ser humano.

- 1) “Iberdrola se lanza con BP a por 12.000 electrolinerías con mil millones”

Expansión Página 3 / 2 de diciembre de 2023

Posibilidad de haber sido generado por colaboración IA/humano: 94%

Posibilidad de haber sido generado íntegramente por IA: 6%

- 2) “Iberdrola lanza su planta de amoníaco en Portugal”

El Economista Página 6 / 12 de enero de 2024

Posibilidad de haber sido generado por colaboración IA/humano: 92%

Posibilidad de haber sido generado íntegramente por IA: 8%

- 3) “Acciona nombra CEO de Infraestructuras a José Díaz-Caneja”

Expansión Página 10 / 15 de diciembre de 2023

Posibilidad de haber sido generado por colaboración IA/humano: 92%

Posibilidad de haber sido generado íntegramente por IA: 8%

La comparación entre estas tres noticias no permite, de nuevo, extraer correlaciones sobre su pertinencia al grupo de contenidos con mayores posibilidades de haber sido generadas humanos sin apoyo de tecnología IA. Por un lado, la extensión de los textos fue heterogénea (355 para Iberdrola en *El Economista*, 327 para Iberdrola en *Expansión* y 394 para Acciona en *Expansión*). Por otro, esas longitudes no estuvieron muy desconectadas de la media de todo el corpus de prensa en papel que fue de 346,5 palabras por noticias.

De la misma forma, la presencia del término entidad protagonista (Iberdrola y Acciona) dentro del texto también fue dispar desde las 3 únicas menciones de Iberdrola en *El Economista*, las 5 de Iberdrola en *Expansión* y las 8 de Acciona en *Expansión*.

Así pues, el análisis de los contenidos no permitió encontrar un elemento clave identificable que pueda predeterminar la detección de mayor participación humana. De igual forma, sí que los análisis por GPTZero demostraron una variable muy evidente y es que tres contenidos digitales nativos, aun estando firmados, tuvieron mayores tasas de participación de IA en comparación con tres contenidos de prensa en papel que tuvieron las mayores tasas de participación humana.

4.5. Correlación número de palabras

Como última variable a tener en cuenta se llevó a cabo un recuento de palabras de los contenidos y se buscó correlaciones entre este dato y el resultado ofrecido por GPTZero segmentando por canal (prensa en papel o digital). Para determinar la existencia o no de vinculación entre ambas variables se aplicó el coeficiente de correlación de Pearson que es una medida estadística que permite entender la relación entre dos conjuntos de datos. De esta forma se pretendió discernir si el volumen de los textos podía condicionar las opciones de que GPTZero detectara en mayor o menor medida la intervención de IA en la producción.

Tabla 4

Recuento del número de palabras en los contenidos analizados por GPZero

<i>Organización</i>	<i>ChatGPT</i>	<i>Cinco Días</i>	<i>Expa nsión</i>	<i>El Econo mista</i>	<i>El Confid encial</i>	<i>OK Diario</i>	<i>Vozpó puli</i>	<i>El Español</i>
<i>Iberdrola</i>	355	327	343	355	814	774	474	611
<i>Naturgy</i>	349	372	376	349	656	594	605	552
<i>Acciona</i>	364	394	360	243	303	557	773	498

Fuente: Elaboración propia.

El análisis discriminó entre noticias publicadas en medios de prensa en papel y publicadas en digitales nativos. En medios digitales no se encontró correlación (el volumen textual no afecta el porcentaje de detección de intervención de IA).

En cambio, en prensa en papel se encontró una correlación débil entre posibilidades de que el texto fuera generado completamente por humanos y el número de palabras (valor de coeficiente 0,3087 cuando el valor para la correlación perfecta sería 1). Esto quiere decir que, cuantas más palabras contenga el texto más opciones hay que GPTZero detecte (acertadamente) de que es un contenido elaborado por humanos (en este caso periodistas profesionales). Sin embargo, como se ha indicado, el valor obtenido (0,3087) no es suficientemente alto para garantizar este extremo ya que en estadística una confiabilidad aceptable sucede en valores de 0,5 o superiores.

Sin embargo, este hallazgo, pese a insuficiente, podría corroborar parcialmente el dato observado anteriormente según el cual los contenidos de prensa en papel tienen mayores posibilidades de ser reconocidos como material creado por humanos. En todo caso podría afirmarse que existe una predisposición a la detección de la participación humana en las noticias de prensa en papel, especialmente en los casos en los que el redactado es proporcionalmente más extenso que sus pares.

Para confirmar este extremo se llevó a cabo la prueba *t* de Student para observar el valor estadístico de la diferencia de medias entre dos grupos. Los grupos que se analizaron fueron el número de palabras de cada texto según la fuente (ChatGPT, *Cinco Días*, *Expansión*, *El Economista*, *El Confidencial*, *OK Diario*, *Vozpópuli* y *El Español*) y el resultado obtenido en GPTZero en cada uno de esos textos.

Los resultados obtenidos fueron los siguientes: ChatGPT 74,8 / *Cinco Días* 17,9 / *Expansión* 35,9 / *El Economista* 8,3 / *El Confidencial* 3,8 / *OK Diario* 9,4 / *Vozpópuli* 6,9 y *El Español* 16,5.

Aunque el caso menos acusado es el corpus de *El Confidencial* se puede concluir que hay una diferencia significativa entre las medias de los dos grupos (número de palabras / probabilidad según GPTZero). Esto es debido a que al comparar los valores promedio de dos conjuntos de datos, encontramos que estos promedios son suficientemente distintos como para concluir que esta diferencia no es una mera coincidencia. Este dato confirma indirectamente con el obtenido en el coeficiente de correlación de Pearson según el cual no puede afirmarse que exista una vinculación de causa / efecto entre el número de palabras o la tipología de los textos en relación con las posibilidades de que estos sean detectados como material presumiblemente generado por AI. Si bien, de la misma forma, no defiende que la prensa en papel tenga unas características que diferencien sus opciones de ser detectada o no como material creado por IA.

5. CONCLUSIONES

Este trabajo ha pretendido abordar una cuestión de primer orden desde la entrada de la IA en la producción de noticias: ¿Existen modelos que pueden detectar eficientemente el uso de esta tecnología en el material elaborado por los medios de comunicación?

La justificación de este propósito se encuentra imbricada en la afectación negativa y la desconfianza que puede generar la IA en los públicos de los medios de comunicación. Por un lado, la IA generativa es una herramienta muy efectiva para la generación masiva de noticias falsas (*fake news*). Por otro, las audiencias receptoras han mostrado sus dudas sobre el uso de IA por parte de productores de noticias cuando estos han sido transparentes sobre su uso, aunque fuera como apoyo al trabajo humano (Longoni *et al.*, 2022)

Para cumplir con este objetivo se llevó a cabo un experimento mediante la que es actualmente la plataforma canónica para detección de IA generativa en textos: GPTZero. Se prepararon tres corpus de textos (generados por ChatGPT, publicados en prensa en papel y publicados en medios digitales nativos) y se analizaron con la plataforma (un análisis con algoritmo pre 17 de enero de 2024 y otro con el algoritmo posterior).

Los resultados fueron llamativos en el sentido de que la detección de IA en los textos elaborados por ChatGPT fue muy baja con el primer algoritmo y todavía peores en el segundo. Es decir, que en el primer análisis la probabilidad de que el texto fuera

reconocido como generado por IA fue baja (media del 7,3%) sobre todo considerando que este dato estuvo no muy alejado de los parámetros obtenidos en textos de medios de comunicación (media del 5,76%).

En el segundo algoritmo los resultados fueron, si cabe, todavía peores ya que el resultado de probabilidad de que los textos generados por ChatGPT fuera, efectivamente, resultado de una IA dieron una media 85,3%. Un resultado elevado que se convierte en una señal de alarma una vez contextualizado, ya que en medios digitales nativos firmados por periodistas se alcanzó una cota mínima en tres casos del 81%. Por tanto, GPTZero en su versión más actualizada no es una garantía para la discriminación entre material generado por IA y material generado presumiblemente por humanos (al señalar falsos positivos).

Esta cuestión (la presunta ineficacia de GPTZero) conduce, sin embargo, a otros hallazgos de interés. El más importante de estos es que con el algoritmo nuevo tres noticias de prensa en papel ofrecieron los resultados más altos de participación humana mientras que otras tres noticias de prensa digital nativa mostraron (como ya se ha comentado) los resultados más bajos.

Un resultado de este tipo puede suponer un peligro añadido para los productores de noticias en medios digitales ya que, aunque no se encontraron rasgos distintivos, podría representar un sesgo según el cual la tipología del canal predetermina de alguna forma las probabilidades de que sus publicaciones sean más proclives a ser marcadas como material con participación de IA. No obstante, este extremo debería confirmarse en investigaciones con corpus más extensos.

Otro hallazgo relevante, pese a tener un peso relativo, observó que analizando estadísticamente la correlación entre número de palabras y resultados de GPTZero (coeficiente de Pearson) había cierta vinculación entre la extensión del redactado de las noticias de prensa en papel y la probabilidad de obtener un valor alto de participación humana según el algoritmo. El resultado concreto ofreció una correlación débil (valor 0,3087 cuando una correlación fuerte sucede con un valor 0,5 o superior) según la cual, cuantas más palabras tiene la noticia de prensa en papel, mayor valor de participación de un ser humano obtiene según ChatGPT. Aunque la prueba *t* para calcular la diferencia de medias no permitió detectar diferencias significativas este dato conlleva a reflexionar si el peso inherente de un medio tradicional como es la prensa publicada en formato físico va más allá del mero halo de prestigio y es una ventaja competitiva frente a productos digitales, aunque estos estén firmados, de la misma forma, por profesionales del sector.

Finalmente, es esencial comprender que la primera premisa de esta investigación no ha podido corroborarse. El modelo actual para la detección de participación de IA en textos (GPTZero) no garantiza la misión para la que ha sido diseñada ya que no pudo determinar como tal, tres noticias creadas por ChatGPT con una simple instrucción. Además, este algoritmo, mostró falsos positivos en colaboración humana/IA (variable *mixed*) en todas las noticias de los corpus textuales, cuando la implantación de la IA en los medios no es todavía lo suficientemente intensa para que sea razonable que todas

ellas tuvieran implicación de IA (especialmente cuando periodistas profesionales firmaron los artículos).

Los resultados de esta investigación no solo no han permitido tranquilizar a los públicos y a los medios de comunicación, sino que despiertan nuevas dudas sobre si el uso masivo de modelos de detección de IA puede ocasionar, en la misma proporción, preocupaciones injustificadas entre los públicos receptores que terminen lesionando el prestigio y el rigor del que son depositarios los medios de comunicación.

6. REFERENCIAS

Aramburú Moncada, L. G., López-Redondo, I. y López Hidalgo, A. (2023). Inteligencia artificial en RTVE al servicio de la España vacía. Proyecto de cobertura informativa con redacción automatizada para las elecciones municipales de 2023. *Revista Latina de Comunicación Social*, 81, 1-16. <https://doi.org/10.4185/RLCS-2023-1550>

Bardin, L. (1991). *Análisis de contenido* (Vol. 89). Ediciones Akal.

Botha, J. y Pieterse, H. (2020). Fake news and deepfakes: A dangerous threat for 21st century information security. En B. K. Payne y H. Wu (Eds.), *ICCWS 2020 15th International Conference on Cyber Warfare and Security* (pp. 57-66). Academic Conferences and Publishing International Limited. <http://tinyurl.com/3zzbpnte>

Deepak, P. (2023). AI and Fake News: Unpacking the Relationships within the Media Ecosystem. *Communication & Journalism Research*, 12(1), 15-32. <http://tinyurl.com/mr3k3d5a>

DeVerna, M. R., Yan, H. Y., Yang, K.-C. y Menczer, F. (2023). Fact-checking information generated by a large language model can decrease news discernment. *arXiv*. <https://doi.org/10.48550/arXiv.2308.10800>

Dunne, M. (2023). Data Dive: Fake news in the age of AI. *IPSOS*. <http://tinyurl.com/nbwfzvu9>

Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methaodos. revista de ciencias sociales*, 11(2), 1-17. <https://doi.org/10.17502/mrcs.v11i2.710>

GPTZero (s.f.) *More than an AI detector. Preserve what's human*. Recuperado [14 de enero de 2025] de <http://tinyurl.com/y7jsnz6s>

de-Lima-Santos, M. F. y Salaverría, R. (2021). From data journalism to artificial intelligence: challenges faced by La Nación in implementing computer vision in news reporting. *Palabra Clave*, 24(3). <https://doi.org/10.5294/pacla.2021.24.3.7>

Longoni, C., Fradkin, A., Cian, L. y Pennycook, G. (2022, June). News from generative artificial intelligence is believed less. En *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)* (pp. 97-106). Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533077>

López del Castillo Wilderbeek, Francisco Leslie.

Inteligencia artificial generativa y confianza en los medios. Un análisis de la detección de IA en noticias usando GPTzero.

- Neuendorf, K. A. (2017). *The content analysis guidebook*. SAGE Publications, Inc. <https://doi.org/10.4135/9781071802878>
- Newman, N. (2023). Journalism, media, and technology trends and predictions 2023. Reuters Insitute for the Study of Journalism. <https://doi.org/10.5287/bodleian:NokooZeEP>
- PR Noticias (s.f.) *Ranking: Prensa*. Recuperado el 24 de octubre de 2023 de <http://tinyurl.com/57u3vzaw>
- Rodríguez-Pérez, C., Seibt, T., Magallón-Rosa, R., Paniagua-Rojano, F. J. y Chacón-Peinado, S. (2023). Purposes, principles, and difficulties of fact-checking in Ibero-America: Journalists' perceptions. *Journalism Practice*, 17(10), 2159-2177. <https://doi.org/10.1080/17512786.2022.2124434>
- Sádaba-Chalezquer, C. y Salaverría-Aliaga, R. (2022). *I Estudio sobre la desinformación en España. Un proyecto de UTECA y la Universidad de Navarra*. <http://tinyurl.com/9t5cxb62>
- Sandrini, L. y Somogyi, R. (2023). Generative AI and deceptive news consumption. *Economics Letters*, 232, 111317. <https://doi.org/10.1016/j.econlet.2023.111317>
- Shoaib, M. R., Wang, Z., Ahvanooey, M. T. y Zhao, J. (2023). Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. En *2023 International Conference on Computer and Applications (ICCA)* (pp. 1-7). IEEE. <https://doi.org/10.1109/icca59364.2023.10401723>

AUTOR:

Francisco Leslie López del Castillo Wilderbeek: Doctor en comunicación por la Universitat Pompeu Fabra, Profesor colaborador en la UOC, Profesor asociado en la UOC. Ejerce como analista de medios de comunicación en la empresa Rebold. Su producción científica reciente se ha centrado en el origen y efectos de la Inteligencia Artificial Generativa (GenAI) en artículos como: "Inteligencia Artificial Generativa: determinismo tecnológico o artefacto construido socialmente" (*Palabra Clave*). franciscoleslie@alumni.upf.edu

Orcid ID: <https://orcid.org/0000-0002-6664-7849>

Google Scholar: <https://scholar.google.es/citations?user=UFyl1AkAAAAJ&hl=es&oi=ao>

ARTÍCULOS RELACIONADOS:

- Abdullah, N. H., Hassan, I., Azura Tuan Zaki, T. S., Ahmad, M. F., Hassan, N. A., Mohd Zahari, A. S., Ismail, M. M. y Azmi, N. J. (2022). Examining the Relationship Between Factors Influencing Political Information Seeking-Behaviour through Social Media among Youths in Malaysia. *Revista de Comunicación de la SEECI*, 55, 1-15. <https://doi.org/10.15198/seeci.2022.55.e746>
- Encinillas García, M. y Martín Sabarís, R. (2023). Desinformación y Salud en la era PRECOVID: Una revisión sistemática. *Revista de Comunicación y Salud*, 13, 1-15. <https://doi.org/10.35669/rcys.2023.13.e312>
- Gómez-Diago, G. (2022). Perspectivas para abordar la inteligencia artificial en la enseñanza de periodismo. Una revisión de experiencias investigadoras y docentes. *Revista Latina de Comunicación Social*, 80, 29-46. <https://doi.org/10.4185/RLCS-2022-1542>
- Martínez-Sánchez, J. A. (2022). Prevención de la difusión de fake news y bulos durante la pandemia de COVID-19 en España. De la penalización al impulso de la alfabetización informacional. *Revista de Ciencias de la Comunicación e Información*, 27, 15-32. <https://doi.org/10.35742/rcci.2022.27.e236>
- Pinto, H., Nogueira, C. y Vieira, G. (2023). Digitalisation landscape in the European Union: Statistical insights for a Digital Transformation. *European Public & Social Innovation Review*, 8(1), 20-38. <https://pub.sinnergiak.org/esir/article/view/233>

VivatAcademia

revista de comunicación

ISSN: 1575-2844