


Evolución del uso de lenguajes de programación y herramientas digitales en el periodismo de datos español

Enrique Ribera-CarbonellUniversidad Miguel Hernández de Elche (UMH) ✉ **Félix Arias-Robles**Universidad Miguel Hernández de Elche (UMH) ✉ <https://dx.doi.org/10.5209/emp.96796>

Recibido: 30 de junio de 2024 / Aceptado: 31 de octubre de 2024

ES Resumen. El periodismo de datos es una de las especialidades con mayor proyección en los medios de comunicación, pero también una de las que mayores exigencias técnicas conlleva. Los lenguajes de programación y las herramientas digitales adquieren aquí un protagonismo especial e implican mayores esfuerzos en la formación de los periodistas. Para analizar la evolución del uso de la programación y de las herramientas digitales en España, se ha llevado a cabo un trabajo de campo longitudinal en dos etapas. Primero, se entrevistó a un profesional de cada medio de comunicación que contara con un equipo consolidado de periodismo de datos. Se consiguieron diez entrevistas en 2021. Este proceso se repitió en 2024 con un total de ocho profesionales de ocho medios para comprobar su evolución durante esos tres años. Los resultados de esta investigación muestran que R, Python y JavaScript son los lenguajes más utilizados en las redacciones españolas para la realización de las piezas periodísticas con datos, mientras que Excel, Google Spreadsheets, Datawrapper y Flourish son las herramientas mejor valoradas por los periodistas para la obtención, limpieza, análisis y visualización de los datos. Por otra parte, se observa además que la inteligencia artificial es una tecnología en auge que, aunque los periodistas utilizan con cautela, se emplea principalmente en la creación y curación de código, al no conseguir resultados totalmente fiables en las pruebas que han realizado para el análisis de los datos.

Palabras clave: Periodismo de datos, lenguajes de programación, herramientas digitales, profesión periodística, inteligencia artificial.

ENG Evolution of the use of programming languages and digital Tools in Spanish data journalism

Abstract. Data journalism is one of the specialties with the greatest potential in the media, but also one of the most technically demanding. Programming languages and digital tools play a special role here and require greater efforts in training journalists. To analyze the evolution of the use of programming and digital tools in Spain, a longitudinal field study was carried out in two stages. First, a professional from each media outlet with a well-established data journalism team was interviewed, resulting in a total of 10 interviews in 2021. This process was repeated in 2024 with a total of 8 professionals from 8 media outlets to observe their evolution over those three years. The results of this research show that R, Python, and JavaScript are the most widely used languages in Spanish newsrooms for producing data-based journalistic pieces, while Excel, Google Spreadsheets, Datawrapper, and Flourish are the tools most valued by journalists for data acquisition, cleaning, analysis, and visualization. Additionally, it was observed that AI is an emerging technology, used cautiously by journalists, mainly for code creation and curation, as it has not yet produced fully reliable results in the data analysis tests they have conducted.

Keywords: Data journalism, programming languages, digital tools, journalistic profession, artificial intelligence.

Cómo citar: Ribera-Carbonell, E. y Arias-Robles, F. (2025). Evolución del uso de lenguajes de programación y herramientas digitales en el periodismo de datos español. *Estudios sobre el Mensaje Periodístico*, 31(1), 253-267. <https://dx.doi.org/10.5209/emp.96796>

1. Introducción

El periodismo de datos no es una disciplina nueva, pero desde 2010 ha ido creciendo hasta convertirse en uno de los pilares de los medios de comunicación (Carrasco-Polaino y Flores-Vivar, 2020). Esta especialidad experimentó un notable auge tras las primeras filtraciones de Wikileaks; a raíz de declaraciones como las de Tim Berners-Lee en las que aseguraba que el futuro de la profesión periodística residía en el análisis numérico, o con la publicación del especial de *The Economist* titulado «Datos, datos en todas partes», donde se bautizaba como el «nuevo petróleo» (Camaj *et al.*, 2022).

El periodismo de datos permite obtener, procesar y visualizar de forma automatizada y atractiva grandes cantidades de información (Vállez y Codina, 2018) sobre cualquier temática gracias a su transversalidad (Córdoba-Cabús *et al.*, 2023). Estos avances han obligado a los periodistas a adaptarse y a formarse para poder dar valor añadido a sus historias (Mair *et al.*, 2015). La evolución de esta especialidad obliga también a fomentar la capacidad crítica entre los profesionales sobre la naturaleza y la repercusión de sus trabajos (Gray y Bounegru, 2021). Y entre los nuevos conocimientos que los profesionales deben adquirir, los lenguajes de programación y el manejo de las herramientas digitales adquiere un especial protagonismo.

Profundizar en cuáles son y cómo se utilizan estos recursos constituye el principal objetivo de esta investigación, que se desarrolla a partir de las siguientes preguntas de investigación (PI):

- PI1. ¿Cuáles son los lenguajes de programación más utilizados en los equipos de datos de los principales medios de comunicación? ¿Qué librerías utilizan para cada lenguaje?
- PI2. ¿Qué herramientas se emplean como alternativas a los lenguajes de programación?
- PI3. ¿Cuáles son las herramientas de inteligencia artificial (IA) más utilizadas? ¿Para qué proceso es más efectiva la IA?

2. La tecnología en el periodismo de datos

La evolución tecnológica hizo posible rastrear y analizar bases de datos con el fin de contextualizar hechos informativos. En ese momento, comienza a hablarse de Database Journalism (DBJ), Computer Assisted Journalism (CAJ) o Computer Assisted Reporting (CAR) (Bounegru *et al.*, 2012). En gran medida, se trataba de derivaciones del ya denominado periodismo de precisión (Meyer, 1973, p. 6), pero con la introducción de ordenadores para examinar grandes cantidades de datos y así descubrir asociaciones o correlaciones estadísticas en todo tipo de documentos (Ferrerías, 2016).

Aunque no haya una diferencia clara entre los objetivos y técnicas del periodismo de precisión y el de datos (Dader, 2002; Ferrerías, 2016), la aparición de esta última etiqueta, a principios del siglo XXI, indica una nueva fase en esta especialidad. El volumen de datos que está disponible en línea, en combinación con sofisticadas herramientas centradas en el usuario, la autopublicación y herramientas *crowdsourcing*, permitía a más profesionales trabajar más fácilmente con mayores volúmenes de información (Bounegru *et al.*, 2012).

El *big data* nace del interés por las grandes cantidades de datos disponibles en internet gracias al uso masivo de los denominados medios sociales (Facebook, Twitter, blogs, etc.) y de otras fuentes textuales de información (medios de comunicación *online*, webs oficiales, etc.) (Broussard, 2019). Se ha intentado vincular el *big data* solo con el tamaño de los datos (por ejemplo, en los papeles de Panamá se usaron 2,6 terabytes). Esta dimensión no es suficiente para caracterizarlo, ya que el concepto se refiere fundamentalmente a volúmenes masivos y complejos de información tanto estructurada como no estructurada, que es recogida durante cierto periodo de tiempo y que requiere de métodos computacionales para extraer conocimiento (Arcila *et al.*, 2016). Otros conceptos importantes ligados al estudio del *big data* también aluden a su intencionalidad y utilidad (Murphy y Barton; 2014).

Por tanto, se puede calificar el periodismo de datos como una disciplina que se manifiesta en tres fases clave del periodismo: el acceso, el tratamiento y la publicación de la información. Y en todas, la tecnología es el denominador común. Probablemente fue Bradshaw (2011) quien mejor conceptualizó la especialidad mediante un diagrama, con forma de pirámide invertida, para representar las fases del proceso de producción. El autor diferenciaba entre cinco niveles: compilar, limpiar, contextualizar, combinar y comunicar. La esencia de esta propuesta se conserva en las versiones posteriores (Antón-Bravo, 2013, p. 112; Ferrerías, 2016, p. 270; La Rosa y Sandoval-Martín, 2016, p. 1212; López López, 2019, p. 111; Lorenzi Reno, 2016, p. 176). El mayor esfuerzo por aunar las características de esta especialización fue probablemente el de Coddington (2015), quien la concebía como la práctica periodística de obtener, reportear, organizar, editar y publicar datos de interés públicos con la aplicación de técnicas estadísticas, computacionales, de visualización y de diseño.

La delimitación de funciones y perfiles de los profesionales de la información vuelven a ser revisada, como en anteriores ocasiones, ante la evolución tecnológica (Sánchez-García *et al.*, 2015). En el periodismo de datos, se necesita un profesional que puede estar formado en documentación, periodismo o cualquier otra disciplina, que posee un aditivo de conocimiento necesario para la gestión de datos y clasificación de información, así como para la búsqueda, procesamiento, interpretaciones concluyentes, análisis de datos y capacidad para crear y publicar información con una narrativa propia que se desprende del propio género utilizado. Deberá manejar grandes cantidades de información, clasificar y después introducir esos datos en las aplicaciones tecnológicas. Además, deben tener conocimientos sobre estadística, programación, diseño, sociología, psicología, además del periodismo (Flores y Cebrián, 2012). Es por este motivo por el que los profesionales de esta disciplina desean recibir más formación sobre el manejo de datos y el ámbito técnico que las relacionadas con el periodismo y la ética (Bisiani *et al.*, 2023).

La literatura especializada coincide en que la naturaleza de esta especialidad obliga a profundizar en dos aspectos clave: la tecnología y el manejo de datos. En relación con el primero, Appelgren y Lindén (2020, p. 61) destacan dos claves: la capacidad para

programar y para diseñar. López-García *et al.* (2016, p. 292) señalan la necesidad de dominar herramientas informáticas, concretamente las relacionadas con la gestión de bases de datos, visualización de información y narrativas multimedia. Broussard y Boss (2018, p. 1213) añadieron más detalles y resaltaron la importancia de conocer *frameworks* y lenguajes de programación como Javascript para crear proyectos. Boyles y Meisinger (2020, p. 189), por su parte, incidían en la necesidad de aprender a identificar las oportunidades y también los riesgos que ofrece la comunicación con las máquinas.

Las habilidades técnicas servirían para las fases finales del proceso, como la visualización de datos o el diseño de interfaces, pero también para las iniciales, como la obtención automatizada de bases de datos (Boyles y Meyer, 2017, p. 435). En esta línea, Engebretsen *et al.* (2019, p. 3) resaltaron la relación entre el desarrollo de fenómenos como la visualización de datos y el auge de la innovación tecnológica. Loosen *et al.* (2017) concluyeron que las habilidades híbridas entre el programador y el periodista de datos actualmente están mejorando el periodismo. Finalmente, Diakopoulos (2020, p. 2) destacó los beneficios del uso de algoritmos para orientar la atención editorial hacia eventos o información potencialmente de interés periodístico antes de la publicación.

Sobre el otro aspecto clave, el tratamiento de datos, Appelgren y Lindén (2020, p. 61) enumeraron la necesidad de conocer los métodos de periodismo de investigación, el conocimiento en estadística, la gestión de datos y la elaboración de informes estadísticos. En este sentido, los periodistas con múltiples habilidades están siendo cada vez más valorados. Davies (2018, p. 110) incidió en la importancia de acabar con la ansiedad por las matemáticas y acercar la utilidad de estos recursos para el periodismo diario. En esta línea, Weiss y Retis-Rivas (2018, p. 3) defendieron la necesidad de acabar con la idea de que las matemáticas y la estadística se alejan de la labor periodística. Sobre todo porque, ahora que la simplificación de las herramientas elimina algunas de las barreras tecnológicas, los editores cada vez le dan más importancia a la capacidad para encontrar temas entre grandes volúmenes de datos (Arias-Robles y López-López, 2020, p. 10).

Para el periodismo de datos, por lo tanto, hace falta entrenamiento y aprendizaje constantes y conocimientos de materias generalmente alejadas de las preferencias del periodista convencional. Se debe conocer el funcionamiento de la administración pública, dominio del inglés para entender las herramientas más utilizadas, ya que no tienen su versión en español, habilidades en búsquedas en la web más avanzadas de las habituales, conocer los principios básicos de programas de análisis de datos (como Excel o Google Spreadsheets), conocimientos de matemática básica y de estadística para poder entrevistar a los números, así como de visualización (Arias-Robles y López-López, 2020; Crucianelli, 2013).

Lewis *et al.* (2020) resaltaron algunas de las claves para la formación en periodismo de datos: dominar la estadística descriptiva y la interpretación de datos, respetar la ética y desplegar el pensamiento computacional. En esta línea, López-García *et al.* (2017, p. 81) resaltaron el incremento de la influencia

de la tecnología en los periodistas de esta especialidad. Baack (2018) ya había avanzado la idea de la hibridación entre periodistas de datos y «tecnólogos sociales», algo en lo que profundizaron Dodds *et al.* (2024) para analizar la complejidad que promueve la colaboración entre diferentes redacciones, facilitada gracias a plataformas como GitHub.

Y todo esto no hace sino crecer ante el auge de la inteligencia artificial (IA), cuya vinculación con el periodismo de datos es indudable. La IA, entendida como «a collection of ideas, technologies, and techniques that relate to a computer system's capacity to perform tasks normally requiring human intelligence» (Beckett, 2019, p. 15), comprende tecnologías como el análisis estadístico o el procesamiento del lenguaje, cuyos elementos han sido parte integral del periodismo desde sus inicios (Deuze y Beckett, 2022, p. 1915). Una investigación de Parratt-Fernández *et al.* (2021) reflejó que los temas más estudiados en las investigaciones sobre IA son la aplicación al periodismo de datos, la redacción por robots y la verificación de noticias.

Esta tecnología lleva años realizando tareas cotidianas en nuestros trabajos sin que le prestáramos atención, pero en la actualidad, con la revolución provocada con la aparición de ChatGPT (Pavlik, 2023, p. 92), se han creado muchas herramientas que permiten realizar tareas complejas en poco tiempo. Estos recursos están entrando en las redacciones de una forma ágil pero cautelosa. A pesar de no ser conscientes ni autónomas, plataformas como ChatGPT consiguen procesar y presentar información de manera que puede parecer humana. Destacan además por ofrecer una mayor accesibilidad y menores requisitos técnicos que las anteriores aplicaciones de IA, lo que ha supuesto un salto para las redacciones (Beckett y Yaseen, 2023, p. 8). Y su margen de desarrollo parece aún enorme (Newman, 2023, p. 35).

Casi dos décadas después de los primeros experimentos con algoritmos de generación automática de lenguaje natural en medios estadounidenses (Zheng *et al.*, 2018, p. 267), la integración de la IA en el periodismo ha alcanzado ya una fase de mayor madurez. Varios grupos mediáticos españoles implantaron proyectos de diversa índole y magnitud incluso antes de la irrupción de ChatGPT (Túñez-López *et al.*, 2021, p. 178). Atrás han quedado las reacciones más extremas, tanto de entusiasmo como de rechazo. En cambio, sí parece consolidarse la idea de que la IA va a determinar el futuro desarrollo de la transformación de los medios (Zheng *et al.*, 2018, p. 273).

No obstante, la emergencia de los grandes modelos de lenguaje ha propulsado avances previamente inimaginables (Maslej, 2023). Se espera que su evolución persista y su impacto en la cotidianidad se intensifique (Chui, 2022). Rao y Verweij (2022) calculaban que la mitad de la integración de nuevas tecnologías en los medios de comunicación se realizará en el corto plazo (menos de tres años). El último informe de JournalismAI (Beckett y Yaseen, 2023) refleja que más del 75 % de los profesionales incorporan la IA en al menos una fase de la cadena de valor periodística, pero que solo una tercera parte de las entidades ha formulado una estrategia para capitalizar esta tecnología.

3. Metodología

Esta investigación se basa en la realización de un estudio de campo longitudinal en dos etapas basado en entrevistas semiestructuradas a periodistas de datos de medios de comunicación. Las primeras entrevistas se realizaron durante la primera mitad de 2021 y las entrevistas comparativas se han realizado durante la primera mitad de 2024 con la finalidad de descubrir la evolución del uso de la programación y de herramientas alternativas, además de identificar los usos que se le da en la actualidad a la IA en las redacciones de medios de comunicación españoles.

La investigación sobre estudio de casos constituye una metodología muy extendida en las ciencias sociales (Jiménez Chaves y Comet Weiler, 2016) para analizar el «cómo» y el «porqué», dos elementos fundamentales en este trabajo sobre el uso de códigos de programación y herramientas tecnológicas (Yin, 2014). Esta estrategia de investigación, que puede basarse en uno o más casos, se enfoca en entender las dinámicas presentes en el fenómeno a explicar y combinan métodos de colección de datos como entrevistas o cuestionarios (Eisenhardt, 1989).

La selección de los medios es fundamental para comprender cómo se está utilizando la programación en las secciones de periodismo de datos de los medios de comunicación en España. El propósito de esta selección es identificar a aquellos profesionales que trabajen, y a ser posible lideren, medios con equipos consolidados y que cuenten con una presencia significativa en el campo, lo que nos permite analizar de manera precisa el grado de integración y aplicación de herramientas de programación en las dinámicas periodísticas actuales. La primera oleada de entrevistas se basó en la audiencia y la repercusión en redes sociales que tuvieron en los dos primeros meses de 2021 sus medios. Entre ellos, la premisa fue que dispusiera de un equipo consolidado dedicado al periodismo de datos.

Para identificar estos medios, se utilizó una matriz que consideraba tanto el número de periodistas de datos —extraídos del proyecto «Periodistas de datos», que recoge una lista de profesionales del periodismo de datos en España y Latinoamérica, elaborado por Félix Arias y Miguel Carvajal—, como su presencia en redes sociales (véase el Anexo 1), lo que permitió identificar 11 medios adecuados para el estudio (*El País*, *Maldita*, *Civio*, *El Confidencial*, RTVE, *Newtral*, *Storydata*, *El Mundo*, *elDiario.es*, *La Sexta* y *Datadista*), al contar todos ellos con varios especialistas en la plantilla y tener relevancia en redes sociales. A continuación, se escogió a un periodista que trabajase en el medio: siempre que fue posible, al responsable de la sección de datos de cada medio. Finalmente, se pudo realizar diez entrevistas por la falta de respuesta de un medio (*El Mundo*).

Para conservar el anonimato de los entrevistados, no se recogen sus nombres en esta investigación, pero su perfil profesional es el siguiente: seis son hombres y cuatro, mujeres; nueve tienen formación en periodismo y uno en ingeniería informática; tres de ellos son cofundadores del medio al que pertenecen, cuatro están a cargo de la sección de datos y los tres restantes forman parte del equipo especializado en datos. Para contactar con los profesionales, se recurrió al correo personal o corporativo, si era público,

y en caso contrario a redes sociales como Twitter o LinkedIn.

Las 16 preguntas se dividieron en dos bloques. En el primero (3) se buscó conocer las opiniones sobre los lenguajes de programación en general y en las situaciones en las que recurren a ellos (PI1). El segundo bloque (7) se centró en descubrir qué lenguajes de programación utilizan para cada parte del proceso periodístico y las librerías que utilizan (PI1), así como las herramientas elegidas como alternativa a estos lenguajes (PI2). Para finalizar, se realizaron seis preguntas de tipo Escala de Likert en las que los entrevistados otorgaron una nota (de 0 a 10) a cada una de las afirmaciones propuestas.

Las entrevistas tuvieron una duración media de 35 minutos. La mayoría se realizaron a través de la plataforma de videollamada Google Meet para poder compartir pantalla para la explicación de las respuestas. En caso de no poder realizarla mediante esta plataforma, se llevaron a cabo vía telefónica con la misma duración media. Para comprobar la idoneidad de las preguntas y la duración de las entrevistas, se realizó un pretest con uno de los seleccionados (*Datadista*). Se comprobó la necesidad de aclarar que las preguntas buscan conocer los pensamientos de los profesionales y la forma en la que trabajan. De estas respuestas se extrajo la información concreta para realizar la muestra cuantitativa y las explicaciones de estas se utilizaron para añadir conocimiento profesional en el apartado de «Discusión y conclusión».

En la selección de los medios para la segunda entrevista, se contactó —a través del correo electrónico corporativo o, en su defecto, mediante redes sociales personales— a los mismos profesionales de los medios que participaron en la primera fase. En los casos en que no se obtuvo respuesta o no pudieron participar, se sustituyeron por otros medios que también tienen equipos consolidados de periodismo de datos, siguiendo los mismos criterios y metodología empleados en la primera entrevista. De esta forma, se consiguió realizar ocho entrevistas a periodistas de *El País*, *Maldita*, *El Confidencial*, RTVE, *Storydata*, *Relevo*, *El Orden Mundial* y *elDiario.es*. Este enfoque permitió mantener la continuidad y comparabilidad de los datos, asegurando la participación de medios con una base sólida en esta especialidad.

Las entrevistas se desarrollaron con el mismo sistema que en la ocasión anterior y con parecida duración media. Las preguntas de estas entrevistas se dividieron en tres bloques, más un último bloque con preguntas del tipo escala de Likert. Los dos primeros bloques han contenido las mismas preguntas que las realizadas en 2021. Se añadió un tercer bloque centrado en el uso de la IA (PI3) en el periodismo de datos. Un único entrevistado tuvo que contestar por escrito, de manera asincrónica, al no disponer de tiempo para atender la entrevista vía telefónica.

4. Metodología

4.1. Lenguajes de programación

Los resultados obtenidos a través de las entrevistas realizadas (véase la Tabla 1) muestran cómo ha evolucionado la forma de trabajar el periodismo de datos en los medios españoles. R se ha convertido en el lenguaje de programación preferido por delante

de Python. Las herramientas alternativas al lenguaje de programación más utilizadas son las mismas en ambos periodos: Datawrapper, Flourish, Excel y

Google Spreadsheets. Por su parte, OpenRefine ha bajado en la frecuencia de uso en estos últimos años.

Tabla 1. Mejores y más utilizados lenguajes de programación para crear contenido periodístico.

2021						2024					
Mejores	N	%	Más utilizados	N	%	Mejores	N	%	Más utilizados	N	%
Python	8	80	JS	6	60	R↑	7	88	R↑	7	88
JS	7	70	R	4	40	Python↓	5	63	JS	6	75
R	6	60	Python	3	30	JS↓	5	63	Python	3	38
Ruby	1	10	Ruby	1	10						
PHP	1	10	Ninguno	1	10						
SQL	1	10									

Fuente: elaboración propia. Las flechas (↑ y ↓) representan una variación de más del 10 %.

Python ha dejado de ser el lenguaje más utilizado por los profesionales entrevistados (80 % en 2021 por un 63 % en 2024). En su lugar, han elegido R (60 % en 2021 por 88 % en 2024). JavaScript sigue estando bien valorado, pero su única utilización en el proceso de visualización ha provocado que se le valore en menor medida (70 % por 63 %). Los entrevistados siguen añadiendo que no hay uno mejor, sino que en función al lenguaje que has aprendido te pa-

rece mejor uno u otro. Del mismo modo que en la primera tanda de entrevistas, se sigue teniendo en cuenta que R es una alternativa muy buena por la amplia comunidad que tiene ya que esta actualiza constantemente. También se ha mencionado que R tiene una interfaz y una forma de tratar los datos que se asemeja más a la hoja de cálculo tradicional, por lo que la entrada y su curva de aprendizaje son más sencillas.

Tabla 2. Motivos a favor y en contra de usar lenguajes de programación.

2021					
A favor	N	%	En contra	N	%
Existencia de grandes cantidades de datos	7	70	Requiere mucho tiempo	4	40
Automatización del proceso	4	40	Escasez de datos	3	30
Extracción de información	3	30	Herramientas digitales más sencillas	1	10
Simplificación de tareas	2	20	Ausencia de datos de calidad	1	10
Creación de visualizaciones especiales	2	20			
Capacidad para cruzar bases de datos	1	10			
2024					
Todo tipo de piezas ↑	3	38	Escasez de datos↑	5	63
Existencia de grandes cantidades de datos↓	3	38	Sencillez del trabajo	3	38
Creación de visualizaciones especiales↑	3	38	Requiere mucho tiempo↓	3	38
Necesidad de buscar tendencias	1	14	Detalle	2	25
Automatización del proceso↓	1	14			
Extracción de información↓	1	14			

Fuente: elaboración propia.

En la actualidad, existe una mayoría de profesionales que defienden la utilización de los lenguajes de programación para todo tipo de piezas (38 %), según muestra la Tabla 2. Por norma general, se siguen apoyando en los lenguajes de programación cuando se

encuentran con grandes cantidades de datos, aunque con menor porcentaje (del 70 % al 38 %). En la primera investigación se adujeron los siguientes motivos: la necesidad de automatizar los procesos (40 %), la necesidad de extraer la información (30 %), la compleji-

dad de las tareas (20 %) y la búsqueda de visualizaciones especiales (20 %). Sin embargo, en la actualidad ha aumentado la valoración en el uso de programación para la creación de visualizaciones especiales (38 %) y ha bajado la automatización del proceso, incluyendo aquí tanto los trabajos recurrentes como la extracción de información (14 % ambas). La necesidad de identificar tendencias ha surgido como una respuesta en esta investigación, a diferencia de lo observado en la realizada en 2021 (14 %). La bajada de contestaciones a favor de trabajos que necesitan de automatización de los procesos o de grandes cantidades de datos tiene sentido al existir medios que han pasado de recurrir a la programación solamente para esas tareas concretas a utilizarlas para todo tipo de piezas. Por este motivo, no significa que se haya dejado de utilizar la programación para estos procesos.

Por otra parte, los periodistas siguen valorando el tiempo del que disponen para negar la utilización de los lenguajes de programación, pero valoran más la escasa cantidad de datos necesarios. Por último, han dejado de mencionar las veces que puedes realizar la misma tarea con otras herramientas y cuando los datos obtenidos están bien estructurados y han pasado valorar los trabajos que requieren de un detalle dato por dato o tareas sencillas que no requieren de mucho análisis. El lenguaje más utilizado por los medios ha dejado de ser JavaScript, aunque ha incrementado su porcentaje de utilización, a ser R. Python sigue siendo el tercero. En la investigación realizada hace tres años se mencionó la utilización de Ruby (10 % por 0 %) o incluso la no utilización de ninguno, aspecto que no ha pasado en la actualidad (10 % por 0 %).

Tabla 3. Librerías utilizadas.

2021											
Librerías JS	N	%*	Librerías R	N	%	Librerías Python	N	%	Librerías Ruby	N	%
D3	9	100	Tidyverse	3	100	Pandas	3	100	Nokogiri	1	100
Vue.js	1	11	Googlesheets	3	100	BeautifulSoup	1	33	Mechanize	1	100
Leaflet	1	11	Countrycode	1	33	Selenium	1	33			
Node.js	1	11	Rmapshaper	1	33	Geopandas	1	33			
			Data.table	1	33	Geocoder	1	33			
						NumPy	1	33			
2024											
Librerías JS	N	%	Librerías R	N	%	Librerías Python	N	%	Librerías Ruby	N	%
D3	6	100	Tidyverse	5	100	Pandas	3	100	—		
Mapbox	2	33	Ggplot	3	60	Selenium↑	2	67			
Geo	1	17	Datawrapper	3	60	BeautifulSoup	1	33			
Svelte	1	17	GeoJson	2	40	Request	1	33			
Hight charts	1	17	Googlesheets↓	2	40						
			Rmapshaper	1	20						
			SimpleFolders	1	20						
			AEMET	1	20						
			Html_table	1	20						
			Googledrive	1	20						
			Janitor	1	20						
			Lubridate	1	20						
			Terra	1	20						

Fuente: elaboración propia.

* Porcentaje de personas que utilizan ese lenguaje y han contestado la pregunta sobre las librerías.

La Tabla 3 muestra que JS es el lenguaje de programación por excelencia para la visualización de datos y su librería estrella sigue siendo D3, que permite hacer visualizaciones interactivas (100 %). Las siguientes librerías en la actualidad son Mapbox, su-

pliendo el uso de Leaflet para crear mapas interactivos, Geo para tratar los datos geográficos. Svelte suple a Vue para construir aplicaciones web y se menciona la librería Hight charts como alternativa a D3 para generar gráficos interactivos. En la actualidad los

medios de comunicación no han mencionado la utilización de Node.js, que es un entorno de ejecución.

En la actualidad se utiliza un mayor número de librerías. Tidyverse (conjunto de librerías diseñadas para la Ciencia de Datos) se mantiene como la librería más empleada por los que trabajan en R (100 %). Google Sheets, librería para conectar R con las hojas de Google sigue manteniéndose como una de las librerías más recurridas, aunque con menos impacto (100 % a 40 %). Caso contrario a la librería de Datawrapper que conecta esta herramienta con R, que ha pasado de no mencionarse a utilizarse por el 60 %. Mismo porcentaje que Ggplot, librería de visualización utilizada para el análisis de datos y no para crear visualizaciones publicables. GeoJson ha sido mencionada por el 40 %, una librería muy valorada para trabajar con datos cartográficos. Con menos uso se encuentran Rmapshaper, que permite editar y simplificar los objetos «geojson», «Spatial» y «sf»; SimpleFolders que facilita la creación y manejo de estructuras de carpetas; AEMET que proporciona acceso a los datos meteorológicos de la Agencia Estatal de Meteorología de España (AEMET); Html_table, permite la extracción de tablas HTML y su conversión a data frames; Google Drive sirve para interactuar con el almacenamiento de Google Drive; Janitor ofrece herramientas para limpiar y examinar datos, especialmente útiles para la limpieza de data frames; Lubridate, que simplifica el manejo y la manipulación de fechas y horas en R y Terra, que proporciona funciones avanzadas para el análisis y manipulación de datos espaciales y raster en R. Countrycode, que sirve para estandarizar los nombres de los países de una base de datos y Data.table, que permite trabajar con grandes bases de datos de una optimizada han dejado de mencionarse en la actualidad.

Todos los que apuestan por Python siguen trabajando con la librería Pandas, una librería dedicada al manejo y análisis de datos. La siguiente librería más utilizada es Selenium, para probar y registrar interacciones en páginas web. Las otras librerías a las que más recurren los periodistas son: BeautifulSoup, para el raspado de páginas web; Request, muy útil para hacer solicitudes HTTP, utilizado para interactuar con APIs y descargar contenido web y Time, que ofrece funciones para manejar y manipular el tiempo, incluyendo mediciones de tiempo y retrasos en la ejecución del código.

Geopandas, para trabajar con datos geoespaciales y Geocoder, librería especializada en geocodificación, es decir, conseguir las coordenadas desde una dirección y viceversa, y NumPy, que da soporte para crear vectores y matrices multidimensionales, junto con una gran colección de funciones matemáticas han dejado de mencionarse en la actualidad.

El único medio que trabajaba con Ruby en 2021 no se le ha podido entrevistar en 2024 y consideraba que las librerías esenciales en esa época eran Nokogiri, que sirve para leer, escribir, modificar y consultar documentos XML y HTML, y Mechanize, que se utiliza para automatizar la interacción con sitios web.

4.2. Herramientas digitales

No se han apreciado cambios significativos en el uso de herramientas digitales de 2021 con la actualidad. La obtención de datos es la fase en la que más herramientas digitales emplean los periodistas (nueve en ambas investigaciones) pese a que la mayoría de periodistas afirman que la mayoría de datos que obtienen es a través de «botones de descarga».

Tabla 4. Herramientas alternativas más usadas en cada uno de los procesos.

Fase del proceso (2021)											
Obtención			Limpieza			Análisis			Visualización		
Herramienta	Nº	%	Herramienta	Nº	%	Herramienta	Nº	%	Herramienta	Nº	%
Tabula	3	38	OpenRefine	7	70	Excel	8	80	Datawrapper	9	100
OpenRefine	2	25	Excel	6	60	Google Sheet	6	60	Flourish	8	89
import.io	2	25	Google Sheets	5	50	Datawrapper	1	10	Ilustrator	4	44
Table Capture	2	25	Calc	1	10	Flourish	1	10	Mapbox	4	44
Web Scraper	1	13				Tableau	1	10	Qgis	2	22
Adobe PDF	1	13				Calc	1	10	Infogram	1	11
Google Sheet	1	13							uMap	1	11
Dataminer	1	13							RAWGraphs	1	11
Abyy FineR	1	13									
Fase del proceso (2024)											
Obtención			Limpieza			Análisis de datos			Visualización de datos		
Herramienta	Nº	%	Herramienta	Nº	%	Herramienta	Nº	%	Herramienta	Nº	%
Table Capture	2	29	Excel	4	57	Excel↓	4	57	Datawrapper	7	100
import.io	1	14	Google Sheet	4	57	Google Sheet	4	57	FLOURISH↓	5	71
Octoparse	1	14	OpenRefine↓	2	29	Open Refine	2	29	QGIS	2	29

Obtención			Limpieza			Análisis de datos			Visualización de datos		
Herramienta	Nº	%	Herramienta	Nº	%	Herramienta	Nº	%	Herramienta	Nº	%
Abyy Fine R	1	14	PowerBi	1	14	Datawrapper	1	14	ILUSTRATOR	2	29
Tabula	1	14				Flourish	1	14	RAWGRAPHS	1	14
WebScraper	1	14				PowerB	1	14			
Adobe PDF	1	14									
PoweBi	1	14									
InstDataScrap	1	14									

Fuente: elaboración propia.

Para obtener datos, tienen a su disposición un gran abanico de herramientas como Tabula, Import.io, Table Capture, WebScraper, Adobe PDF o Abbyy FineReader, todas estas mencionadas desde 2021. Otras herramientas mencionadas en la actualidad pero que no se tenían en cuenta en el pasado son: Octoparse, PowerBi e Instant Data Scraper. Todo lo opuesto que OpenRefine, Google Spreadsheets o Dataminer, que han dejado de mencionarse en las entrevistas de 2024.

A la hora de limpiar datos, se ha dejado de apostar por OpenRefine (70 % en 21 a 29 % en 2024), y se ha optado por un uso mayoritario de Excel y Google Spreadsheets. Excel se utiliza más para ciertas operaciones no admitidas en Google Spreadsheets y esta se utiliza para trabajos colaborativos. Otra herramienta que se utiliza es Power Bi, el hermano avanzado de Excel en Microsoft. No se ha mencionado el uso de Calc, la hoja de cálculo de Libreoffice, en las redacciones consultadas.

En el análisis de datos las dos herramientas por excelencia siguen siendo Excel y Google Spreadsheets. Después de estas están ha aparecido en la actualidad OpenRefine, no mencionada esta vez para la obtención pero si para el análisis. Las herramientas Datawrapper y Flourish, a pesar de tener una finalidad de visualizadores de datos, se emplean para buscar tendencias en las gráficas y poder realizar compro-

baciones previas y sacar conclusiones. Otra herramienta a mencionar es Tableau, que ha dejado de utilizarse de hace tres años a la actualidad.

A la hora de visualizar los datos, Datawrapper y Flourish son las favoritas de los periodistas. Datawrapper va por delante por su mayor rapidez de carga y su visualización limpia. Muchos de ellos utilizan Flourish cuando la visualización deseada no se puede realizar en Datawrapper. Esta tendencia se ha acentuado con los años y cada vez menos periodistas recurren a Flourish. Otras herramientas mencionadas son Illustrator, Qgis y RAWGraphs. Los periodistas han pasado de mencionar ocho herramientas a mencionar solamente cinco.

4.3. Inteligencia artificial

Para evaluar de qué manera se integra la IA en las redacciones, se ha empezado este bloque del cuestionario preguntando en qué puede ayudar más esta tecnología al día a día de un periodista de datos. La ayuda en la programación es la tarea en la que más confían los periodistas de datos que la IA les va a poder ayudar. Tanto puliendo código como generando. También se mencionan otras tareas como la automatización de procesos, resúmenes de documentos, asistente virtual, búsqueda de documentación académica o en la limpieza de datos.

Tabla 5. ¿En qué ayuda más la IA a un periodista de datos?

Tareas	Nº	%
Programación	6	75
Automatización	1	13
Resúmenes	1	13
Asistente	1	13
Búsqueda de documentación académica	1	13
Limpieza de datos	1	13

Fuente: elaboración propia.

La percepción mayoritaria (Tabla 6) de los periodistas es que la IA no debe utilizarse 100 % en ninguna tarea ya que tiene varios errores y no es confiable. También mencionan de manera repetida el trabajo editorial y el *fact checking* como tareas que no se deben entregar a la IA. Para uno de los encuestados,

automatizar análisis no debería realizarse nunca con esta tecnología ya que el análisis realizado y el enfoque que se le da a los datos será siempre más cierto si tendrá en cuenta más variables ambientales si lo ejecuta un profesional.

Tabla 6. ¿Hay alguna tarea que nunca se debe realizar mediante IA?

Tareas	Nº	%
Ninguna tarea se debe realizar 100 % mediante IA	3	38
Editorial	2	25
Fact checking	2	25
Automatizar análisis	1	13

Fuente: elaboración propia.

No inventar información y atribuir fuentes son dos de las tareas pendientes que tiene la IA para mejorar según los profesionales de datos de los medios de comunicación españoles, según se observa en la Tabla 7. Tener más transparencia en la ejecución de los procesos, mejorar la programación desde 0, que

sean herramientas gratuitas y accesibles, mejorar la generación de gráficos y la integración con otras herramientas como RStudio son otras de las tareas a mejorar para recurrir a la IA en las redacciones de manera continuada.

Tabla 7. ¿En qué puede mejorar la IA?

Tareas	Nº	%
No inventar	3	38
Atribuir fuentes	2	25
Transparencia	1	13
Programar desde 0	1	13
Gratuito	1	13
Mejorar gráficos	1	13
Integración con otras herramientas	1	13

Fuente: elaboración propia.

En cuanto a las herramientas más utilizadas, destaca de forma clara ChatGPT, puesto que es utilizada por todos los periodistas consultados en este estudio. Las demás quedan ya a una distancia considerable. Midjourney es la otra herramienta utilizada por más de un periodista: uno de ellos la utiliza para ob-

tener los datos de PDF borrosos o de poca calidad y otro para generar imágenes que acompañan al reportaje. Ninguna de las demás herramientas, como se aprecia en la Tabla 8, es citada por más de un periodista entrevistado (de modo que no se supera el 13 %).

Tabla 8. ¿Qué herramienta de IA utilizáis más?

Tareas	Nº	%
ChatGPT	8	100
Midjourney	2	25
Notion	1	13
BlackBox	1	13
Jasper	1	13
Research Rabbit App	1	13

Fuente: elaboración propia.

A pesar de ser ChatGPT la herramienta más mencionada, en el trabajo diario la mayoría de los entrevistados dice no utilizar ninguna herramienta de IA en la actualidad por culpa de los errores que hemos comentado anteriormente. Los periodistas que si se

apoyan de estas herramientas, sobre todo utilizan herramientas para la visualización de datos. En la obtención de datos se utiliza ChatGPT, Midjourney para la limpieza de PDF y Apify para scrapear páginas web. En la limpieza solo utilizan ChatGPT, normal-

mente para detectar nombres o repeticiones. Para el análisis, además de ChatGPT, se utiliza Julius, herramienta de IA enfocada al análisis de datos avanzado. Para la visualización se utilizan herramientas más va-

riadas como las mencionadas ChatGPT o Midjourney, Aino.word para la creación de mapas o gráficas a partir de texto plano y Supadash, que permite generar gráficas en pocos segundos.

Tabla 9. ¿Qué herramienta de IA utilizas en cada parte del proceso?

Herramienta	Obtención		Limpieza		Análisis		Visualización	
	Nº	%	Nº	%	Nº	%	Nº	%
Ninguna	6	75	5	63	6	75	5	63
ChatGPT	1	13	3	38	2	25	1	13
Midjourney	1	13					1	13
Apify	1	13						
Julius					1	13		
Aino.world							1	13
Supadash							1	13

Fuente: elaboración propia.

4.3. Visión general sobre la programación

Para conocer la visión general sobre programación entre los periodistas de datos se realizaron una serie de afirmaciones que tuvieron que puntuar del 0 al 10, siendo 0 en desacuerdo y 10 totalmente de acuerdo. La pregunta «La programación es un recurso muy utilizado para la creación de contenidos periodísticos» ha sido generalmente valorada por los entrevistados de forma neutra en ambas ocasiones, salvo por un medio que le otorgó la puntuación más alta (10) y otro que le asignó un 1, obteniendo una media y mediana de 5,5, una moda de 5 y una desviación típica de 2,3. En la actualidad la situación no ha variado, pues las respuestas son neutras, excepto un 8 y un 2, que han sido las únicas respuestas más extremas y las que han provocado que la media y mediana baje a 5, la moda suba a 6 y la desviación típica baje a 1,4.

En la pregunta «Todo periodista debería saber programación» ha habido una gran diversidad de respuestas en ambas ocasiones. Dos medios la valoraron con un 0 mientras que tres medios le pusieron un 8 en la encuesta de 2021, consiguiendo una media de 3,7, una mediana de 2,5, una moda de 8 y una desviación típica de 3,4. En 2024 solo un medio la ha valorado con la peor respuesta y sí ha existido un medio que ha valorado con un 10 a esta pregunta ocasionando una subida en la media a 5,6, y en la mediana a 6, una bajada en la moda a 3 y una desviación típica de 3,6. Deja clara la diferencia de criterios en esta afirmación. La pregunta «La programación potencia la creación de contenidos periodísticos» contó con el mayor apoyo por parte de los especialistas, con una media de 8,9 en 2021 y 8,6 en 2024, una mediana de 9,5 en la primera tanda y de 8,5 en la segunda, una moda de 10 que ha bajado a 8,5 en la actualidad y una desviación típica de 1,6 por 1,1 en 2024.

En el cuestionario también se han incluido afirmaciones sobre la IA basadas en el estudio previo y la experiencia de los investigadores con estas herramientas. En la afirmación «La IA es fundamental para

resolver dudas de programación» ya se pudo comprobar la tendencia de las puntuaciones en las preguntas previas, pero muchos bajaron la puntuación que le iban a asignar por pensar que no es fundamental al existir otras plataformas que pueden resolver estas dudas, pero si lo ven muy eficaz y facilita encontrar respuesta exacta al problema y no tener que adaptar las respuestas parecidas de foros. En este sentido, la media fue de 6,13, la mediana de 6,5, una moda de 6 y una desviación típica de 2,5.

Las puntuaciones de la afirmación «La IA ayuda mucho a pulir código» han sido acordes a lo comentado a lo largo del cuestionario. Consiguen una media y mediana de 7,5, una moda de 8 y una desviación típica de 1,6. La última afirmación «La IA es muy útil para descubrir fórmulas para hojas de cálculo» también fue puntuada con notas altas. Reciben una media de 7,5, una mediana de 8, una moda de 9 y una desviación típica de 2,2.

5. Discusión y conclusión

Los resultados muestran que el uso de la programación se va asentando en el día a día de las redacciones españolas y que este aumento lleva a la estandarización de su uso y prácticamente se limite a los lenguajes de programación R, Python y JS. El código, como sucede también más allá de Occidente, se afianza en España como un catalizador de la innovación en el periodismo de datos (Dodds *et al.*, 2024, p.10). La investigación también refleja que el uso de la programación potencia la creación de todo tipo de proyectos periodísticos, especialmente los que cuentan con un gran volumen de datos y permiten mostrar estos datos de diversas formas innovadoras y únicas que diferencian a los medios de comunicación. Y esto no ha hecho más que crecer con la implantación de la IA (Túñez-López *et al.*, 2021, p. 184).

Por otra parte, las entrevistas con responsables de cinco empresas tecnológicas que colaboran con medios de comunicación han permitido entender aspectos importantes sobre el proceso de implan-

tación de la IA periodística en España: en particular, sobre los motivos (laborales, profesionales, tecnológicos o incluso psicológicos) por los que hasta ahora no se han aprovechado por completo las posibilidades que brindan a los periodistas estas herramientas.

Todos los medios encuestados utilizan programación para elaborar sus piezas. Esto se puede traducir en una mejor preparación a la hora de cubrir sucesos especiales como fue la pandemia, en la que muchos profesionales invirtieron tiempo para poder aprender programación y trabajar con bases de datos muy pesadas o extraer grandes cantidades de información de las páginas web mediante el *web scraping* (raspado). Las herramientas digitales se siguen utilizando para los temas de los que no disponen de mucho tiempo o poco volumen de datos, por lo que trabajar de forma híbrida con programación y herramientas permite a los medios publicar informaciones más profundas y otras más ligeras optimizando el tiempo.

Entre los lenguajes de programación más utilizados (PI1), R se ha convertido en el más utilizado gracias a su potente comunidad que permite resolver dudas y errores rápidamente. JavaScript, gracias a su potente librería de visualización D3, es el lenguaje más utilizado para la visualización al permitir crear gráficos únicos, mientras Python, al que se le valora la versatilidad, se le da un menor uso. En la mayoría de los casos, la selección de librerías en distintos lenguajes de programación está determinada por la necesidad de ejecutar tareas específicas, lo que lleva a los periodistas a buscar y utilizar aquellas librerías que mejor se adapten a esos requisitos. Sin embargo, existen librerías que dan la posibilidad de realizar tareas complejas en la ciencia de datos que las convierten en esenciales para los periodistas. Este es el caso de Tidyverse en R o Pandas en Python presentes en todas las redacciones.

En la actualidad los periodistas utilizan menos las herramientas alternativas. Sin embargo, siguen siendo muy recurridas las hojas de cálculo de Excel o Google Spreadsheets para analizar, sobre todo, bases de datos pequeñas. La elección de una herramienta u otra se basa en si en el medio se trabaja de forma colaborativa (Google Sheets) o si el trabajo es personal (Excel) por lo que la forma de trabajar de cada departamento influye en la elección. En cuanto a la visualización, Datawrapper es la herramienta más utilizada y cuenta con una fluida integración con R y Python, que permite utilizarla simultáneamente con programación. Flourish, sin embargo, se utiliza gracias a su gran cantidad de gráficas diferentes. La prevalencia de estas dos herramientas en los medios de comunicación ha llevado a que muchas de las visualizaciones publicadas por estos medios presenten características similares (PI2). En contraste, el uso de JavaScript para la creación de gráficos permite desarrollar visualizaciones con una identidad única, dado que requiere programación personalizada para cada gráfica, evitando así la homogeneidad observada con herramientas preconfiguradas.

La investigación evidencia que la programación está ganando terreno en las redacciones españolas, permitiendo la creación de proyectos periodísticos innovadores y profundos, como se ha estudiado especialmente en países escandinavos (Dodds *et al.*,

2024, p.10). A medida que se estandarizan lenguajes como R, Python y JavaScript, la calidad de las piezas obtiene una identidad singular. Para potenciar aún más su uso, es esencial fomentar la formación continua en programación entre los periodistas y promover la creación de plantillas estandarizadas que agilicen el uso de la programación en tareas sencillas en la que utilizan herramientas de terceros por falta de tiempo.

La investigación sobre el uso de la inteligencia artificial (IA) en el periodismo de datos en España refleja un panorama de adopción cautelosa. A pesar de que la IA, especialmente herramientas como ChatGPT, ha demostrado su utilidad en la generación y resolución de dudas o errores de código, la confianza de los periodistas en estas tecnologías es limitada (Pavlik, 2023, p. 92). Los datos muestran que la programación es el área donde los periodistas de datos ven más beneficios de la IA, seguida de la automatización de procesos y la asistencia en la limpieza de datos (PI3). Sin embargo, persisten serias reservas sobre la delegación completa de ciertas tareas a la IA.

La adopción de la IA en el periodismo de datos tiene un gran potencial para revolucionar la eficiencia y la precisión en el manejo de grandes volúmenes de información. Sin embargo, para que esta adopción sea más amplia y efectiva, es crucial abordar las limitaciones actuales (Becket y Yaseen, 2023, p. 35). La IA debe avanzar en su capacidad para proporcionar resultados más precisos y confiables, y se deben desarrollar políticas claras sobre su uso ético en el periodismo. Futuras mejoras podrían enfocarse en la capacitación de los periodistas en el uso de estas tecnologías y en la creación de sistemas de IA más transparentes y con mejores mecanismos de verificación de la información. Esto permitirá no solo una mayor confianza en la IA, sino también una integración más efectiva en las prácticas periodísticas diarias.

Son evidentes las limitaciones de esta investigación. Cuenta con una muestra valiosa, aunque pequeña (diez medios entrevistados la primera ronda y ocho la segunda), y se circunscribe a una especialidad muy concreta, como es el periodismo de datos, en un solo país: España. Se excluyen otras secciones de los medios de comunicación que también emplean las técnicas analizadas para la elaboración de sus piezas periodísticas, así como otros países, que permitirían un análisis más profundo del uso de la programación, herramientas e IA en el periodismo. Se abre vía la vía para futuras investigaciones que comparen la metodología de trabajo en diferentes naciones de la Unión Europea o entre territorios hispanohablantes. Queda pendiente, además, el uso de las propias tecnologías estudiadas, principalmente Python y R, para investigar sobre el *big data* en periodismo (Arcila-Calderón *et al.*, 2016, p. 623).

No obstante, la investigación aporta valor especialmente desde el punto de vista descriptivo, centrándonos en un aspecto clave para la innovación en el periodismo como es el uso de lenguajes de programación y herramientas digitales para mejorar el producto informativo y la eficiencia en las tareas profesionales. En este estudio, que abarca 18 entrevistas a periodistas de medios relevantes en dos tandas espaciadas por tres años, se ha explorado cómo

la adopción de tecnologías avanzadas, particularmente en el campo del periodismo de datos, refleja el futuro de la profesión. Esta investigación, centrada en una especialidad puntera, subraya la creciente in-

terconexión entre tecnología y periodismo. A través de esta aportación, proporcionamos una visión crucial sobre cómo la innovación tecnológica está transformando las prácticas periodísticas.

6. Contribución de autores

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

7. Referencias bibliográficas

- Arcila-Calderón, C., Barbosa-Caro, E. y Cabezuelo-Lorenzo, F. (2016). Técnicas big data: análisis de textos a gran escala para la investigación científica y periodística. *Profesional de la información*, 25(4). <https://doi.org/10.3145/epi.2016.jul.12>
- Appelgren, E. y Lindén, C.-G. (2020). Data Journalism as a Service: Digital Native Data Journalism Expertise and Product Development. *Media and Communication*, 8(2), 62. <https://doi.org/10.17645/mac.v8i2.2757>
- Antón Bravo, A. (2013). El periodismo de datos y la web semántica. CIC Cuadernos de Información y Comunicación, 18, 99–116. https://doi.org/http://dx.doi.org/10.5209/rev_CIYC.2013.v18.41718
- Arias-Robles, F. y López López, P. J. (2020). Driving the Closest Information. Local Data Journalism in the UK. *Journalism Practice*, 15(5), 1-13. <https://doi.org/10.1080/17512786.2020.1749109>
- Baack, S. (2018). Practically Engaged. The entanglements between data journalism and civic tech. *Digital Journalism*, 6(6), 676-692. <https://doi.org/10.1080/21670811.2017.1375382>
- Beckett, C. (2019). *New powers, new responsibilities: A global survey of Journalism and Artificial Intelligence*. The London School of Economics.

- Beckett, C. y Yaseen, M. (2023). Generating Change. A global survey of what news organisations are doing with AI. *JournalismAi*. <https://www.journalismai.info/research/2023-generating-change>
- Bisiani, S., Abellan, A., Arias-Robles, F. y García-Avilés, J.A. (2023): The Data Journalism Workforce: Demographics, Skills, Work Practices, and Challenges in the Aftermath of the COVID-19 Pandemic, *Journalism Practice*, 1-21 <https://doi.org/10.1080/17512786.2023.2191866>
- Bounegru, L., Chambers, L. y Gray, J. (2012). *The Data Journalism Handbook 1*. O'Reilly Media. <https://datajournalism.com/read/handbook/one>
- Boyles, J. L. y Meisinger, J. (2020). Automation and Adaptation: Reshaping journalistic labor in the newsroom library. *Convergence*, 26(1), 178-192. <https://doi.org/10.1177/1354856518772542>
- Bradshaw, P. (2011). The inverted pyramid of data journalism. *Online Journalism Blog*. <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>
- Broussard, M. (2019). *Artificial unintelligence: how computers misunderstand the world*. The MIT Press. <https://doi.org/10.1080/1369118x.2019.1576752>
- Broussard, M., y Boss, K. (2018). Saving Data Journalism. *Digital Journalism*, 6(9), 1206-1221. <https://doi.org/10.1080/21670811.2018.1505437>
- Camaj, L., Martin, J., y Lanosga, G. (2022). The Impact of Public Transparency Infrastructure on Data Journalism: A Comparative Analysis between Information-Rich and Information-Poor Countries. *Digital Journalism*. <https://doi.org/10.1080/21670811.2022.2077786>
- Carrasco-Polaino, R. y Flores-Vivar, J. M. (2020). Periodismo de datos y visualización. En Sotelo-González y Martínez-Arias, S. *Periodismo y nuevos medios: perspectivas y retos*, pp. 249-274. Madrid, Gedisa.
- Chui, M. (2022). *The state of AI in 2022—and a half decade in review*. McKinsey. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review#/>
- Coddington, M. (2015). Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331-348. <https://doi.org/10.1080/21670811.2014.976400>
- Córdoba-Cabús, A., Huber, B. y Farias-Battle, P. (2023). Data journalism in Spain and Austria: features, organizational structure, limitations, and future perspectives. *Profesional de la información*, 32(1), e320116. <https://doi.org/10.3145/epi.2023.ene.16>
- Crucianelli, S. (2013). ¿Qué es el periodismo de datos? *Cuadernos de Periodistas*, 26.
- Dader, J. L. (2002). *Periodismo de Precisión. Vía socioinformática de descubrir noticias*. Madrid, Síntesis.
- Davies, K. (2018). Getting Started with Data Journalism: A Baby Steps Approach. *Asia Pacific Media Educator*, 28(1), 106-111. <https://doi.org/10.1177/1326365X18767460>
- Deuze, M. y Beckett, C. (2022). Imagination, algorithms and news: Developing AI literacy for journalism. *Digital Journalism*, 10(10), 1913-1918. <https://doi.org/10.1080/21670811.2022.2119152>
- Diakopoulos, N. (2020). Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journalism. *Digital Journalism*, 8(7), 945-967. <https://doi.org/10.1080/21670811.2020.1736946>
- Dodds, T., Reséndez, V., von Nordheim, G., Araujo, T. y Moeller, J. (2024): Collaborative Coding Cultures: How Journalists Use GitHub as a Trading Zone. *Digital Journalism*, 12(7). <https://doi.org/10.1080/21670811.2024.2342468>
- Engelbrechtsen, M., Kennedy, H., y Weber, W. (2019). Data Visualization in Scandinavian Newsrooms. *Nordicom Review*, 39(2), 3-18. <https://doi.org/10.2478/nor-2018-0007>
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.
- Ferreras Rodríguez, E. M. (2016). El Periodismo de Datos en España. *Estudios sobre el Mensaje Periodístico*, 22(1), 255-272. https://doi.org/10.5209/rev_ESMP.2016.v22.n1.52594
- Flores-Vivar, J. y Cebrián Herreros, M. (2012). El data journalism en la construcción de mashups para medios digitales. *Congreso Periodismo Digital*, 215-230.
- Gray, J., y Bounegru, L. (2021). *The Data Journalism Handbook 2*. Bruselas, European Journalism Centre. <https://datajournalism.com/read/handbook/two>
- Jiménez-Chaves, V. y Comet-Weiler, C (2016). Los estudios de casos como enfoque metodológico. *ACADEMO Revista de Investigación en Ciencias Sociales y Humanidades*, 3(2), 1-11.
- La Rosa, L. y Sandoval-Martín, T. (2016). La insuficiencia de la Ley de Transparencia para el ejercicio del Periodismo de datos en España, 1208-1229. <https://doi.org/10.4185/RLCS-2016-1142>
- Lewis, N. P., McAdams, M. y Stalph, F. (2020). Data Journalism. *Journalism and Mass Communication Educator*, 75(1), 16-21. <https://doi.org/10.1177/1077695820904971>
- Lindgren, C. A. (2021). Writing With Data: A Study of Coding on a Data-Journalism Team. *Written Communication*, 38(1). <https://doi.org/10.1177/0741088320968061>
- Loosen, W., Reimer, J. y De Silva-Schmidt, F. (2017). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013-2016. *Journalism*, 21(9). <https://doi.org/10.1177/1464884917735691>
- López López, P. J. (2018). La enseñanza-aprendizaje del periodismo de datos en las facultades españolas. Estudios y propuestas de viabilidad. [Tesis doctoral, Universidad Complutense de Madrid]. <https://docta.ucm.es/entities/publication/Ofed7a7c-92c7-4cec-beed-609d4b4420b8>
- López-García, X., Toural-Bran, C. y Rodríguez-Vázquez, A.-I. (2016). Software, estadística y gestión de bases de datos en el perfil del periodista de datos. *El Profesional de La Información*, 25(2), 286-294. <https://doi.org/10.3145/epi.2016.mar.16>
- López-García, X., Rodríguez-Vázquez, A. I. y Pereira-Fariña, X. (2017). Competencias tecnológicas y nuevos perfiles profesionales : desafíos del periodismo actual. *Revista Científica de Educomu-*

- nicación, XXV(53), 81–90. <https://doi.org/10.3916/C53-2017-08>
- Mair, J., Radcliffe, D. y Felle, T. (2015). *Data Journalism: Inside the Global Future*. Abrams.
- Maslej, N. (2023). *The AI Index 2023 Annual Report*. New York, Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf
- Meyer, P. (1973). *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Rowman & Littlefield Publishers.
- Murphy, B. M. y Barton, J. (2014). From a Sea of Data to Actionable Insights: Big Data and What It Means for Lawyers. *Intellectual Property y Technology Law Journal*, 26(3), 8-18.
- Newman, N. (2023). *Journalism, Media, and Technology Trends and Predictions 2023*. Reuters Institute. <https://doi.org/10.60625/risj-0s9w-z770>
- Parratt-Fernández, S., Mayoral-Sánchez, J., y Mera-Fernández, M. (2021). Aplicación de la inteligencia artificial al periodismo: análisis de la producción académica. *El Profesional de La Información*, 1–12. <https://doi.org/10.3145/epi.2021.may17>
- Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism and Mass Communication Educator*, 78(1), 84–93. <https://doi.org/10.1177/10776958221149577>
- Rao, A. y Verweij, G. (2022). Sizing the prize. PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution. PwC. <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>
- Sánchez-García, P., Campos-Domínguez, E. y Berrocal Gonzalo, S. (2015). Las funciones inalterables del periodista ante los perfiles multimedia emergentes. *Revista Latina de Comunicación Social*, 70, 187-208. <https://doi.org/10.4185/RLCS-2015-1042>
- Túñez-López, J. M., Fieiras Ceide, C. y Vaz-Álvarez, M. (2021). Impacto de la Inteligencia Artificial en el Periodismo: transformaciones en la empresa, los productos, los contenidos y el perfil profesional (Impact of Artificial Intelligence on Journalism: transformations in the company, products, contents and professional profile). *Communication & Society*, 34(1), 177-193.
- Vállez, M. y Codina, L. (2018). Periodismo computacional: evolución, casos y herramientas. *Profesional de la Información*, 27(4), 759. <https://doi.org/10.3145/epi.2018.jul.05>
- Weiss, A. S., y Retis-Rivas, J. (2018). 'I Don't Like Maths, That's Why I am in Journalism': Journalism Student Perceptions and Myths about Data Journalism. *Asia Pacific Media Educator*, 28(1), 1–15. <https://doi.org/10.1177/1326365X18780418>
- Yin, R. (2014). *Case Study Research: Design and Methods*. Sage Publications.
- Zheng, Y., Zhong, B. y 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>

Enrique Ribera Carbonell. Graduado en Periodismo (2021); Máster Universitario en Innovación en Periodismo (2022) por la Universidad Miguel Hernández de Elche; Máster Propio en Periodismo de Datos y Visualización por la Universidad de Alcalá (2022). Miembro del Departamento de Comunicación y Marketing internacional de COEX International Trade, Spin-Off de la Universidad Miguel Hernández de Elche, especializada en asesoramiento, consultoría y formación en Comercio Exterior e Internacionalización de Empresas, desde agosto de 2021. También ha participado en la autoría de libros y comunicaciones en congresos y en la organización de congresos internacionales organizados por la UMH y la UPV. Participa como ponente en el proyecto formativo Inic[IA] para transferir conocimientos esenciales y prácticos sobre IA a profesionales y ha participado en otros proyectos universitarios como LocalDataLab. ORCID: <https://orcid.org/0000-0002-9599-2561>

Félix Arias Robles. Profesor Titular en la Universidad Miguel Hernández de Elche (UMH). Vicedecano del Grado en Periodismo y subdirector del Máster en Innovación en Periodismo de la UMH. Doctor con mención internacional en Nuevos Modelos Periodísticos por la UMH y Máster en Investigación en Periodismo por la Universidad Complutense de Madrid (UCM). Su investigación ahora se centra en la intersección entre la inteligencia artificial, el periodismo de datos y la información local. Ha publicado más de 60 artículos en revistas de alto impacto. Ha sido investigador invitado en JournalismAI, el *think tank* de la London School of Economics (LSE), la Escuela de Periodismo de la Universidad de Sheffield y la Universidad de Leeds (Reino Unido). Ha capacitado en innovación, herramientas digitales e IA a profesionales de la comunicación de grupos como *La Voz de Galicia*, Vocento, Onda Regional de Murcia, el Cádiz Club de Fútbol o la Asociación de Informadores de Elche. Actualmente lidera el proyecto formativo Inic[IA] para transferir conocimientos esenciales y prácticos sobre IA a profesionales. ORCID: <https://orcid.org/0000-0001-5910-1541>

Anexo 1

Audiencia, relevancia en redes sociales y periodistas de datos en los medios españoles el 25/03/2021

Medio	Matriz	Twitter	Facebook	Instagram	Audiencia	Periodistas
Maldita	Digital	117500	63000	31300	1.266.222	7
El País	Prensa	8000000	5300000	1100000	19.095.008	5
Civio	Digital	44900	10300		83.710	6
El Confidencial	Digital	910900	1000000	142000	19.263.246	5
RTVE	Audiovisual	1200000	778000	263000	9.722.684	5
Newtral	Digital	187700	33500	40300	401.091	5
Storydata	Agencia	2173	156	734		4
El Mundo	Prensa	3900000	2600000	504000	22.732.110	2
eldiario.es	Digital	1200000	576000	153000	10.691.610	3
La sexta	Audiovisual	1000000	479000	123000	9.632.075	2
Datadista	Digital	26200	8200	855		2
La Vanguardia	Prensa	1200000	4300000	534000	34.972.749	0
Telecinco	Audiovisual	1700000	1600000	1000000	10.899.626	0
ABC	Prensa	525500	1600000	363000	20.059.749	0
Antena 3	Audiovisual	1400000	1900000	402000	10.130.797	0
20 Minutos	Prensa	1400000	2900000	197000	14.858	0
El Español	Digital	427700	733000	90700	20.726.488	0
El Periódico	Prensa	696000	823000	95400	12.739.528	0
OkDiario	Digital	331000	868000	164000	10.729.767	0
Cuatro	Audiovisual	1100000	839000	236000	4.214.905	0
La Razón	Prensa	525500	446000	117000	8.426.492	0

Fuente: elaboración propia.