

Generación automática de textos periodísticos



La inteligencia artificial
aplicada a informativos 2019-2020

Este informe se ha desarrollado bajo la investigación “Inteligencia artificial aplicada al periodismo 2019-2020” por la Cátedra RTVE-UAB Innovación en los informativos en la sociedad digital.

Equipo de investigación

Por parte de la UAB: José Manuel Pérez Tornero, Cristina Pulido, Santiago Tejedor, Laura Cervi, Diana Sanjinés, Wei Zhang, Sally Tayie

Por parte de RTVE: José Juan Ruiz, Pere Vila, David Corral, Carmen Pérez, Esteban Crespo, Esteban Mayoral, Miguel Ángel Martín, Pedro Cánovas.

Con la colaboración de

red **INNO** NEWS

CSO2017-90819-REDT



Generación automática de textos periodísticos por [OI2 RTVE-UAB](#) está licenciado bajo [Creative Commons Reconocimiento-NoComercial 4.0 Internacional License](#).

Índice

Introducción.....	4
Marco conceptual	7
Etapas de la escritura automatizada	11
Procesando el lenguaje natural	15
Métodos de la escritura automática.....	17
Método de planificación de historias	18
Estructuras de eventos	18
Redes neuronales.....	20
Perspectivas.....	22
Las primeras aplicaciones	23
El desplazamiento de los periodistas.....	26
Herramientas disponibles.....	28
Heliograf.....	29
Cyborg.....	30
Wordsmith.....	31
Quill	32
Syllabs.....	33
Panels by Getty images and Vizual.ai.....	34
Editor	36
Estudios de caso de escritura automática.....	37
AGENCIA EFE.....	37
DIARIO SPORT.....	44
EL CONFIDENCIAL	50
NARRATIVA	58
Conclusiones	65
Referencias	67



Introducción

Este informe - intitulado **Generación automática de textos periodísticos**- muestra una etapa más en el desarrollo del programa de investigación centrado en la innovación del periodismo audiovisual que Radio Televisión Española (RTVE) y la Universidad Autónoma de Barcelona (UAB) pusieron en marcha en el año 2015. Concretamente es una etapa de la línea de investigación sobre Periodismo e Inteligencia artificial (IA), centrada en estudiar las oportunidades de transformación que ofrecen al periodismo los diversos sistemas basados en el procesamiento inteligente de la información.

El primer informe en la línea de periodismo e IA se dedicó a la detección automática de noticias. Trataba de explorar las posibilidades de disponer de sistemas automáticos de exploración de los contenidos de las redes sociales con objeto de detectar acontecimientos de valor periodístico. Este segundo informe se dedica, en concreto, a analizar las oportunidades que ofrece la generación automática de textos en el campo periodístico.

Pero antes de entrar en el tema específico del informe, recordemos que la línea general de investigación sobre periodismo e IA tiene los siguientes objetivos:

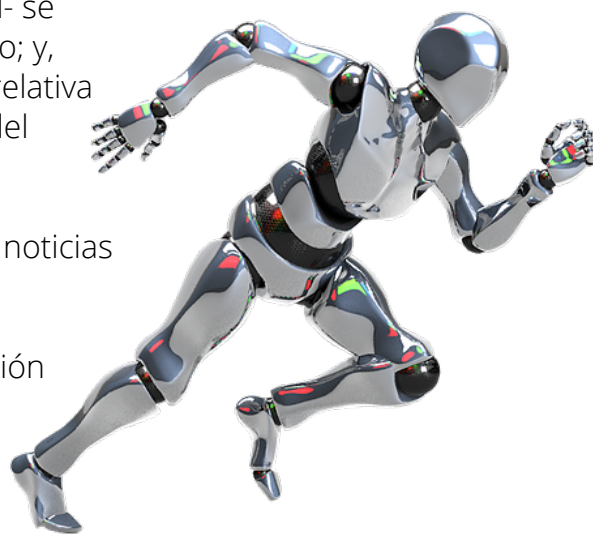
- Conocer, describir y analizar el panorama actual de las diversas aplicaciones de la inteligencia artificial en el periodismo.
- Identificar cómo estas aplicaciones pueden incorporarse, con éxito, a las redacciones audiovisuales.
- Estudiar el impacto, positivo o negativo de la inteligencia artificial en el trabajo periodístico.
- Finalmente, establecer un mapa sobre las tendencias de futuro en el sector que ayude a las tareas de innovación y mejora.

Dentro de estos objetivos, procuramos: a) contar con un mínimo cuadro conceptual que nos permita un adecuado conocimiento del campo de investigación; b) disponer de un análisis de las tendencias del sector; c) identificar las principales herramientas tecnológicas disponibles en el mercado; d) conocer algunos casos concretos de aplicaciones de éxito; y, siempre que sea posible, d) realizar alguna experiencia de aplicación en el marco de RTVE.

En su día definimos como la línea de investigación sobre periodismo e inteligencia artificial partía del reconocimiento de las diversas fases de producción y difusión de la información televisiva en las que la inteligencia artificial puede incidir. Desde el momento que se produce un hecho noticiable, hasta el instante en que este -en forma de texto audiovisual- se publica o se difunde en un determinado canal o medio; y, posteriormente, el tiempo en que dicha información relativa se archiva de cara a su reutilización -tanto por parte del periodista como del público espectador o usuario-.

Este informe dedicado a la generación automática de noticias se organiza en torno a los siguientes temas:

- El conocimiento científico disponible sobre la cuestión
- El desarrollo industrial existente y las aplicaciones comerciales existentes
- Las perspectivas de futuro, tanto desde el punto de vista tecnológico como profesional



- Las consecuencias de la aplicación de la inteligencia artificial en el trabajo periodístico

Trataremos, en primer lugar, de conceptualizar lo que llamamos generación automática de noticias; posteriormente, identificar las herramientas disponibles en el mercado, conocer algunos de los usos que se hacen de ellas en el marco de organizaciones periodísticas; y, finalmente, contribuir a la reflexión sobre las eventuales consecuencias de la aplicación de la inteligencia artificial al campo del periodismo. En este sentido, nos basaremos en la bibliografía científica y académica sobre la materia; a continuación, propondremos un marco conceptual para el proceso de generación automática de textos periodísticos; luego presentaremos los que consideramos los casos más significativos de herramientas tecnológicas comercializadas; y, finalmente, describiremos las tendencias que se viven en nuestro campo, analizando sus repercusiones y consecuencias en el trabajo futuro de los periodistas.





Marco conceptual

¿Qué se entiende por generación automática de textos periodísticos? ¿Cómo participa la inteligencia artificial en esa generación? ¿Qué supone para el trabajo periodístico habitual? ¿Qué consecuencias laborales y éticas puede tener?

No es nuestro objetivo aquí responder a todas estas preguntas, sino sencillamente establecer los términos de la cuestión para poder disponer de un mínimo marco conceptual que nos ayude describir el estado actual de la generación automática de textos en el campo del periodismo.

Empezaremos por tratar de definir que entenderemos por generación automática de textos periodísticos.

La generación automática de los textos periodísticos tiene el objetivo de producir textos de carácter periodístico utilizando el potencial de la informática, los ordenadores, la programación y la inteligencia artificial. Esta producción implica en un momento u otro, la intervención y supervisión de los

periodistas, por lo que en esencia no es nunca enteramente automática.

La intervención humana en este proceso puede ir desde una participación del periodista simplemente inicial –en la ideación y programación del sistema- hasta una intervención continua del periodista –que se inicia con la programación, continúa con el entrenamiento y supervisión del sistema y puede seguir con la posibilidad siempre abierta de intervención.

Por su parte, la actividad autónoma del sistema informático puede ser con poca o sin ninguna intervención humana. En los últimos tiempos, la autonomía de los sistemas de inteligencia artificial no ha parado de crecer. No suelen responder ya, únicamente, a una programación algorítmica previa, sino que son capaces de aprender por sí mismos y de auto programarse en función de ese aprendizaje. Es lo que se conoce como *machine learning*.

Recientemente, **las aplicaciones de la inteligencia artificial a la generación de noticias respondían a motivos de eficacia productiva**. Un sistema debidamente programado podía utilizar bases de datos estructuradas –por ejemplo, datos económicos, de bolsa, o meteorológicos- y trasladarlos a textos, también estructurados, que mostraban parte de esa información.



Por ejemplo, una base de datos sobre temperaturas y humedad relativa del aire en todos los municipios de un país podía ser utilizado por un sistema de generación de textos para generar una noticia correspondiente a cada ciudad. El

...la generación automática de los textos periodísticos tiene el objetivo de producir textos de carácter periodístico...

procedimiento es, en principio sencillo. Se programan de antemano ciertas frases o enunciados bien formados desde el punto de vista periodístico y en ellos se establecen ciertos términos que pueden ser cambiados sistemáticamente por la máquina utilizando las bases de datos a su alcance.

Imaginemos la frase:

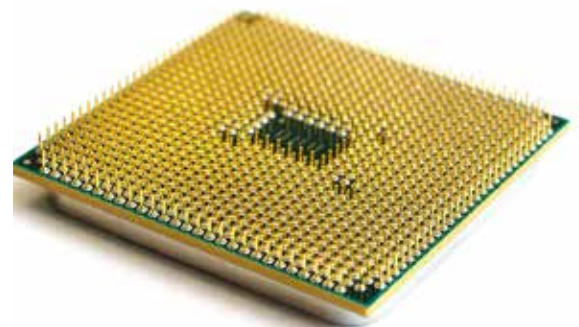
“Mañana en **Grenoble**, la temperatura media será de 15 grados, y la humedad relativa del aire de **50%**”. Simplemente sustituyendo la palabra marcada en negrita por los datos contenidos en las bases disponibles, se pueden generar un texto diferente por cada uno de los municipios de un país, y a una velocidad enorme.



Este simple mecanismo, que, por supuesto, puede sofisticarse ampliamente, tiene la virtud de aumentar enormemente la productividad de textos periodísticos. Con simplemente algunas bases de datos estructurados, y una programación correcta de algoritmos, un determinado sistema puede llegar a producir miles de textos periodísticos aceptables que un ser humano tardaría muchísimo en realizar.

Todo esto ha facilitado que la investigación en IA y periodismo se haya centrado, especialmente, en la **edición automática de noticias**, en inglés *automated news curation* que cubre no solo la generación automática de textos, sino, también, su cuidado y supervisión (Shangyuan y otros, 2019; Carlson, 2015). Sin embargo, conviene hacer algunas aclaraciones.

Para identificar el periodismo que incluye IA en la generación de textos periodísticos se acuñó, en principio, el **término periodismo automatizado** -en inglés, *automated journalism*- que se define como aquel periodismo que **“utiliza algoritmos o software de IA para convertir datos en textos periodísticos con poca o ninguna intervención humana, a excepción de la programación inicial del algoritmo”** (Carlson, 2015: 414). Esta idea también es considerada por otros autores tales como Haim y otros (2017), Monti (2019) y Shangyuan y otros (2019). Este término es el más utilizado en el ámbito académico centrado en la investigación de IA y periodismo (Monti, 2019). Pero existen otros como periodismo computacional, periodismo robot, periodismo algorítmico, máquina de escritura periodística, escritura periodística



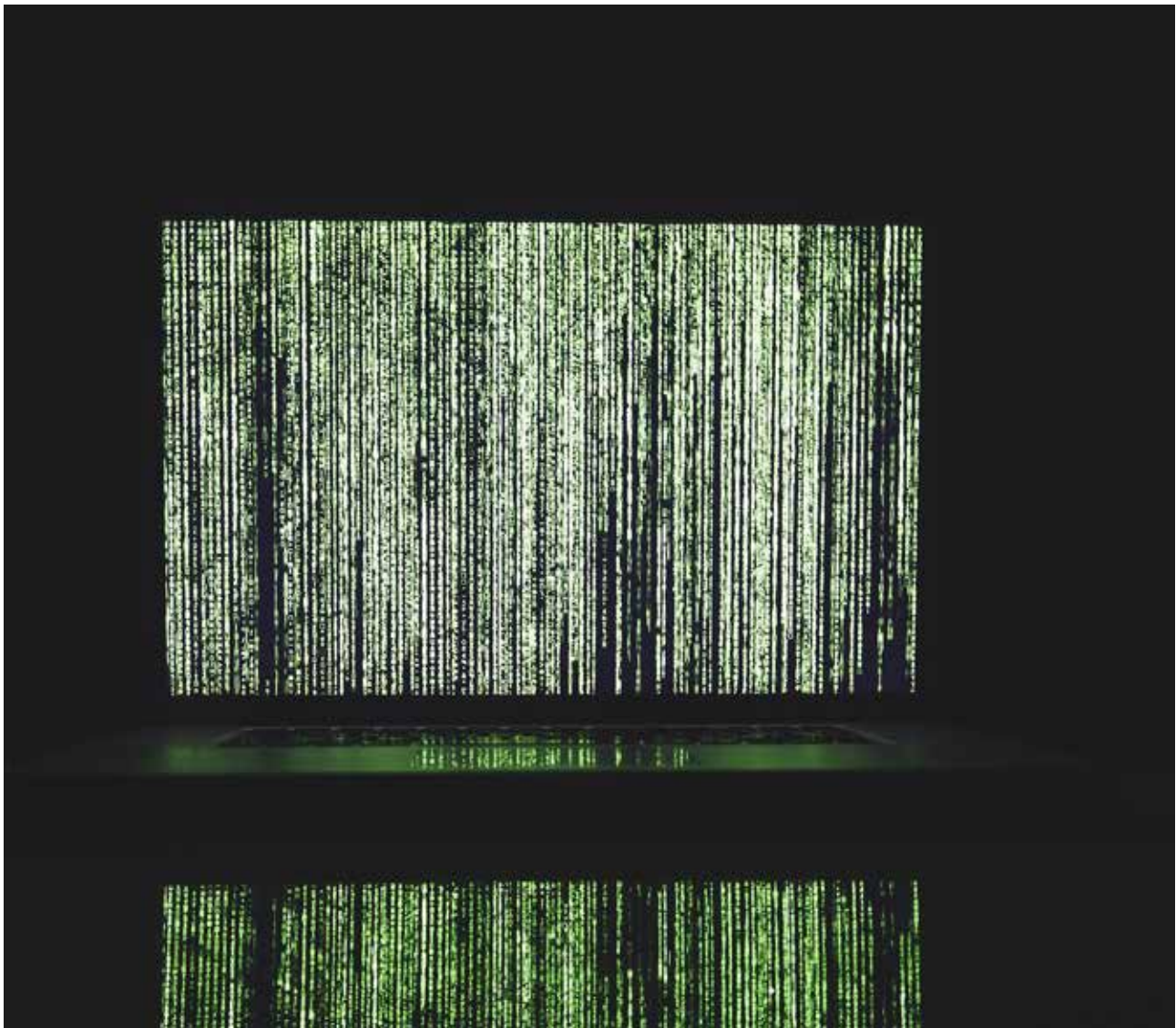
computarizada, periodismo automático, etc. Muchos términos diversos para un campo que también es diverso. Tal vez por ello, conviene poner un poco de orden a partir de algunas consideraciones generales.

La primera y muy general: **no conviene identificar automatización con inteligencia artificial**. De hecho, hay procesos de automatización –incluso de escritura– que no necesitan el concurso de la inteligencia artificial. Esta idea es defendida por Lewis y otros (2019: 417) que señala el aspecto diferencial de la IA: **“la automatización se refiere principalmente a procesos habilitados para IA que incluyen la disponibilidad de datos estructurados y algoritmos capaces de extraer inferencias”**. Siguiendo las indicaciones de Lewis, el periodismo automatizado formaría parte del periodismo computacional y periodismo algorítmico, pero no, en cambio, del ámbito de la inteligencia artificial. Aunque se pueden reconocer diferencias entre la escritura automática y la escritura soportada por la inteligencia artificial, a efectos prácticos, su impacto en el campo periodístico puede ser semejante.

En todo caso, sí queremos excluir del campo de estudio de este trabajo lo que se denomina el uso de robots de noticias –en inglés *news bots*–, o “cuentas automatizadas” que participan en la difusión de noticias e información en las plataformas de redes sociales y que constituyen aproximadamente un 8% de las cuentas en Twitter y Facebook –que, es obvio, por otra parte, muchas empresas, estados, partidos, lobbies, organizaciones y particulares están utilizando para manipular la información (Lokot y otros, 2016)–. Tampoco tratamos aquí en el emergente campo de investigación en comunicación humano-máquina –en inglés, *Human-Machine Communication* (HMC)– donde se investiga sobre la interacción humano-computadora, humano-agente o humano-robot y que está dando lugar a una amplia integración de agentes automáticos en la conversación entre personas y máquinas e incluso en conversaciones entre personas mediadas por máquinas (Lewis y otros, 2019).

Finalmente, preferimos hablar de **generación o escritura automática de noticias (o textos periodísticos) que, de periodismo computacional o automatizado**, porque entendemos que el periodismo es una actividad muy amplia de la que la escritura es sólo una parte. En el periodismo entran procesos de exploración, análisis, comunicación, y toma de decisiones que no están incluidas en el campo que aquí tratamos de estudiar.





Etapas de la escritura automatizada

Las etapas involucradas en el periodismo automatizado no difieren mucho del flujo de trabajo de un periodista tradicional. La diferencia radica en el tipo de actor implicado en la realización de las tareas y etapas.

Si en el periodismo clásico el actor único o principal es el periodista, cuando intervienen los ordenadores o la inteligencia artificial conviene tener muy claro en qué tipo de proceso se implican estos. El siguiente cuadro (ver figura 1) trata de visualizar las etapas en que se puede dividir la labor

periodística en lo que se refiere a las posibilidades que brinda la escritura o edición automática.



Figura 1: Proceso de producción automática de textos periodísticos |

Fuente: Graefe, 2016

La **primera etapa consiste en recolectar datos** de diferentes fuentes, base de datos y/o plataformas (ver Informe 01). Los datos suelen ser, en la mayoría de los casos, estructurados.

En la **segunda etapa, se seleccionan estos datos con criterios** predefinidos o se establecen criterios a partir del tratamiento de analizar esos datos con técnicas de Big Data. De esta tarea, resultará un análisis que permite identificar los datos que resultan más relevantes e interesantes y que son susceptibles de ser incorporados al flujo de información periodística.

En la siguiente etapa, en el caso de que se incorporen computadores o procesos de inteligencia artificial, **es la máquina la que selecciona o determina qué información debe priorizarse** y qué debe filtrarse (Diakopoulos, 2015). Es así como la máquina establece la prioridad o el orden de relevancia de dichos eventos, siguiendo unas reglas predefinidas según criterios periodísticos.

...es la máquina
la que
selecciona o
determina qué
información
debe
priorizarse...

A continuación, en la **cuarta etapa integrando los datos seleccionados** y usando técnicas de procesamiento de lenguaje natural se redactan los textos periodísticos.

Finalmente, en **la última etapa, los textos se difunden** por los canales establecidos de antemano.

En el campo de la escritura automática, las etapas directamente involucradas son las que van desde la segunda etapa a la cuarta. Su participación en la primera no puede ser autónoma. Y consideraremos que en la última no podemos hablar ya de escritura, sino de difusión. Pongamos algunos ejemplos que nos ayuden a comprender la naturaleza de cada una de estas etapas.

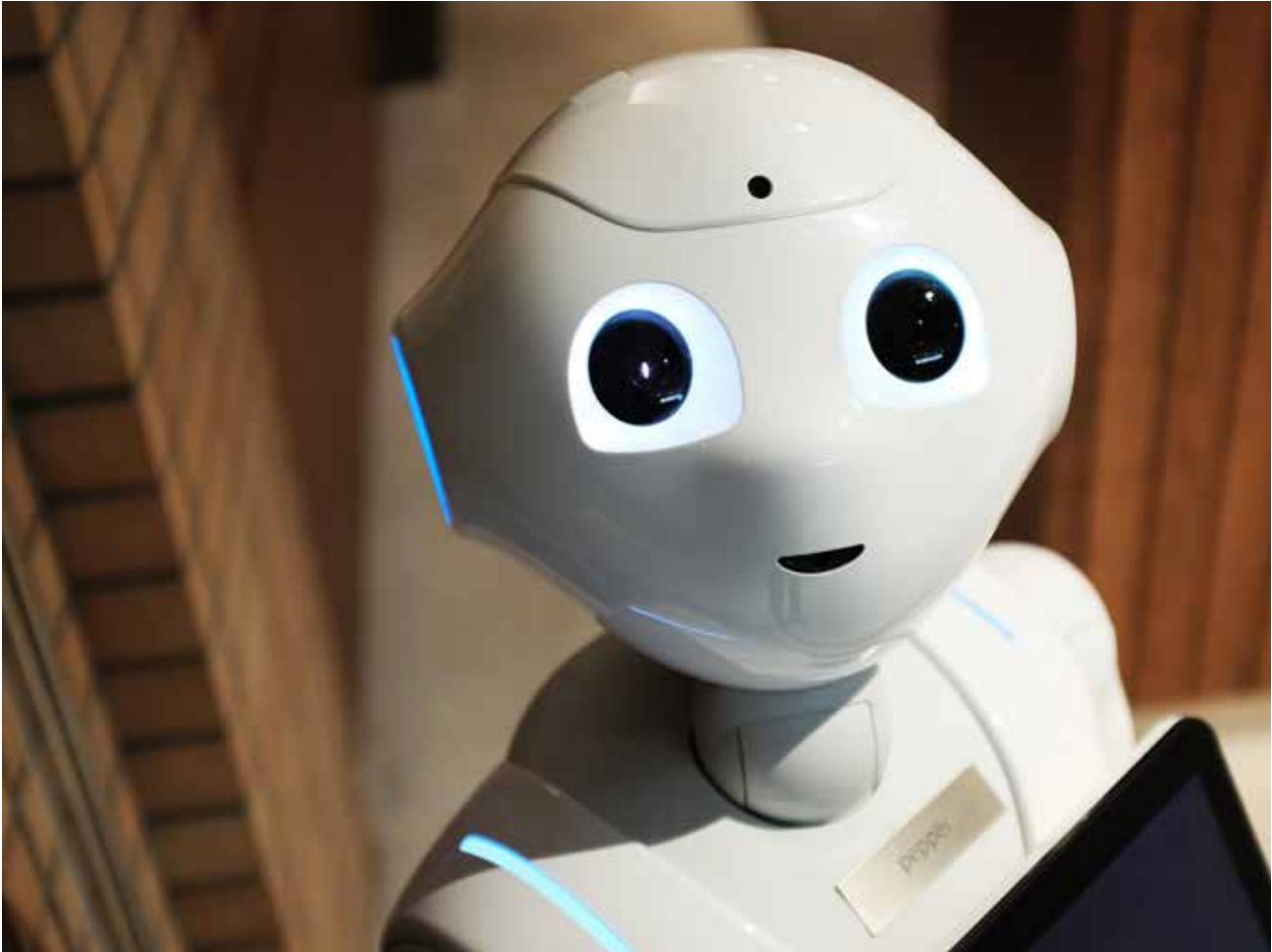
En el caso de noticias sobre fenómenos sísmicos, clima, deportes y de economía, pongamos por caso, existen o se pueden componer bases de datos estructuradas: sobre sismos registrados y su grado; temperaturas, precipitaciones, fenómenos atmosféricos, etc.; resultados de competiciones, marcadores, rankings, etc.; datos sobre valores de bolsa, variaciones de precios, estadísticas de consumo y producción, etc. Pues bien, en todos estos casos, y muchas veces al margen de la labor periodística, existen bases de datos muy organizadas y estructuradas que se mantienen al día y que pueden ser susceptibles de interesar al público de los medios periodísticos. Fueron, de hecho, estas bases las primeras que se utilizaron de cara a facilitar la escritura periodística automatizada (Graefe, 2016). Son además las que han dado lugar a aplicaciones comerciales implantadas en los usos de muchos medios.



Sin embargo, también se pueden automatizar datos no estructurados. Este es el caso del sistema de Thomson Reuters, presentado por Liu y otros (2017). El sistema, llamado *Reuters Tracer*, automatiza la producción de textos usando datos de Twitter (tweets - datos no estructurados, con un máximo de 140 caracteres-) y genera un titular corto para cada hecho o evento. En este caso, no se parte ni de estructuras de datos cerradas, ni de categorías de relevancia preestablecidas, sino de procesos que la máquina desarrolla por sí misma –aprendiendo autónomamente- y que puede dar lugar a contenidos periodísticos. Sin embargo, su valor en la cadena de producción periodística habitual es mucho más azaroso, y no es comparable con el que puede aportar una base de datos estructurada, aunque desde el campo del *Deep learning* se está consiguiendo que no sea tan importante tener datos estructurados o no.

...también se pueden automatizar datos no estructurados...





Procesando el lenguaje natural

Uno de los puntos importantes y complejos para tener en cuenta a la hora de comprender en qué consiste la escritura automática de textos es la noción de generación de lenguaje natural -en inglés *Natural Language Generation* (NLG)-. La generación de lenguaje natural (NLG) se relaciona con otras áreas: a) el procesamiento del lenguaje natural -en inglés *Natural Language Processing* (NLP-), y b) la comprensión del lenguaje natural -en inglés *Natural Language Understanding* (NLU)-. Estas tecnologías son usadas por la IA para procesar y utilizar el lenguaje natural humano (Veel, 2018).

En el caso del procesamiento de lenguaje natural, o PLN, es el campo de la Inteligencia Artificial que da a las máquinas la capacidad de leer, entender y derivar el significado de los lenguajes humanos (ver figura 2). En particular, esta tecnología se utiliza para cuando las máquinas leen el lenguaje y lo convierte en datos estructurados (Nichols, 2017).

En el caso de la comprensión del lenguaje natural, o ULN, es un subtipo de PLN y se utiliza para que las máquinas “comprendan” el lenguaje. Esta tecnología se usa en Alexa y Siri.

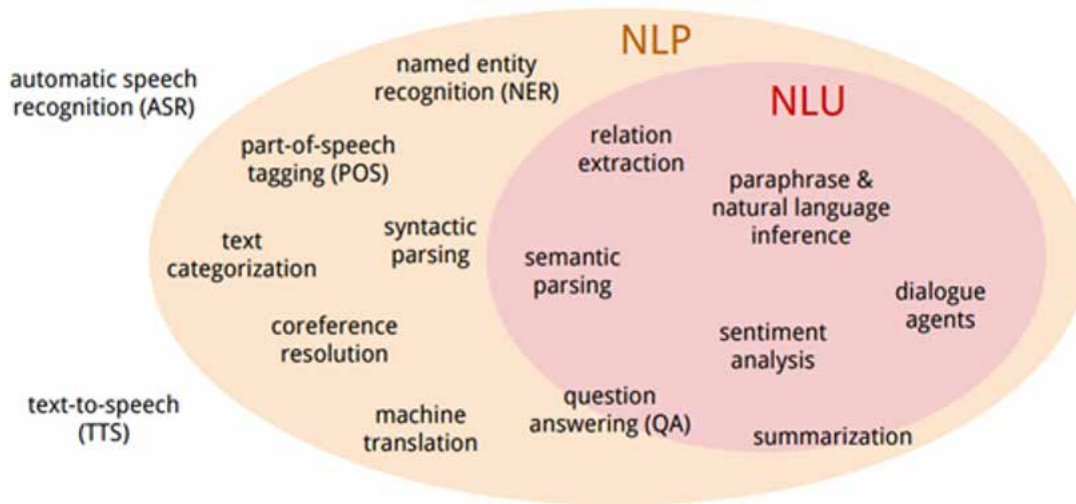


Figura 2. Técnicas de las tecnologías NLP y NLU. | Fuente: Bill MacCartney, 2014.

En el caso de la generación del lenguaje natural, o GLN, se usa cuando las máquinas escriben lenguaje; es decir, se crea texto automáticamente a partir de datos estructurados (Caswell y otros, 2018). Por lo tanto, este informe se centrará en esta tecnología.

Los primeros textos creados con esta tecnología eran textos descriptivos, cortos, de unas secciones/tópicos determinados (como, por ejemplo, deporte, economía, etc.) y de calidad. Sin embargo, se está investigando para crear textos periodísticos más complejos.





Métodos de la escritura automática

Dentro de la generación de lenguaje natural se puede considerar que existen tres métodos: planificación de la historia -*story planning*- (Caswell y otros, 2018), definición de estructuras de eventos para contar historias (*event structures for storytelling*) y generación de historias neuronales (*neural story generation*) (Yao y otros, 2019). Entendiendo el término historia ("story") como un mecanismo de representación del conocimiento para organizar eventos temporales.

Estos eventos son abstracciones generalizadas donde debe estar definido la acción del evento y los roles semánticos de los participantes, entidades, ubicaciones implicadas en ese evento, permitiendo, así, que los datos que varían se sitúen adecuadamente en determinadas secuencias (Caswell y

otros, 2018). Dicho de otro modo, una historia es cualquier información que se suministra ordenadamente a través de la descripción de un conjunto de eventos vinculados causal y lógicamente, que involucran los mismos o personajes (con pocas variaciones), y que suceden en una unidad de tiempo determinada (Yao y otros, 2019).

Método de planificación de historias

Este método se usa desde los años 70 y trata de componer automáticamente una secuencia de eventos que pueden ser contados en forma de historia. En este método se podrían diferenciar tres etapas: a) planificación de documentos; b) micro-planificación (input) y c) realización (output) (Dörr, 2016).

En general, en la primera etapa, planificación de documentos, se identifica la información que es relevante para el usuario, tanto en el input como en el output. En la segunda se establece la relación existente entrada-salida; es decir, entre la planificación de documentos y la realización. En esta etapa, se utilizan algoritmos para seleccionar elementos de un conjunto de datos básico y asignarles relevancia. Finalmente, en la última etapa se decide las palabras a usar y el orden en el que aparecerán.

Estructuras de eventos

En el método de estructuras de eventos (Caswell y otros, 2018) lo esencial es disponer de datos recopilados de modo organizado. A partir de ellos, es posible generar infinidad de tipos de textos que describen o muestran estos datos.

Si se considera que el periodismo narra historias basadas en eventos (hechos) (en inglés, *event-based text narratives*), éstos se deberían codificar y almacenar, en la medida en que estos eventos son la característica central de las noticias. Por esta razón, proponen un prototipo de base de datos (*story database*) donde almacenan eventos (hechos) y textos narrativos. A partir de esta base de datos y de las plantillas que se pueden crear utilizando las herramientas de escritura automática tales como *WordSmith* de *Automated Insights* se generarían los textos periodísticos. Estas plantillas, en realidad, son reglas de lógica y fragmentos de texto asociados (Caswell y otros, 2018).

...una historia es cualquier información que se suministra ordenadamente a través de la descripción de un conjunto de eventos vinculados causal y lógicamente...



Por ejemplo, utilizando disyuntivas concretas:

Si el campo1 es X entonces escribe “texto1” si no [si el campo es Y] escribe “texto2”.

Para cada modelo de datos se debe crear una plantilla diferente, y esta permite ensamblar tales datos con diversos fragmentos de texto. Es posible utilizar cualquier conjunto de datos, pero éstos deben tener siempre el formato requerido en el modelo de datos utilizado en esa plantilla.

...el reto en este tipo de programación es que los textos resultantes tengan coherencia gramatical...

El reto en este tipo de programación es que los textos resultantes tengan coherencia gramatical. Y esta depende de tres factores: a) la calidad de los datos; b) la calidad de los fragmentos de texto utilizados; y c) la coherencia general resultante. De aquí que la creación de estas plantillas se confíe generalmente a periodistas experimentados tanto en el manejo de datos como en redacción periodística. Pero para ello (Caswell y otros, 2018) es necesario que los periodistas desarrollen un estilo de pensamiento computacional, es decir que conozcan el tipo de datos y categorías que son susceptibles de ser codificados, así como que tengan capacidad contrastada a la hora identificar y determinar hechos y textos narrativos como para validarlos y almacenarlos en las plantillas, y poder construir, de este modo, textos complejos (Ver figura 3). Por lo tanto, parece que será necesario: “desarrollar nuevos estilos periodísticos y vocabularios gramaticales apropiados para el periodismo automatizado, y también desarrollar pautas editoriales y guías de estilo para fragmentos de texto, creación de plantillas y ensamblaje de documentos” (Caswell y otros, 2018: 491).

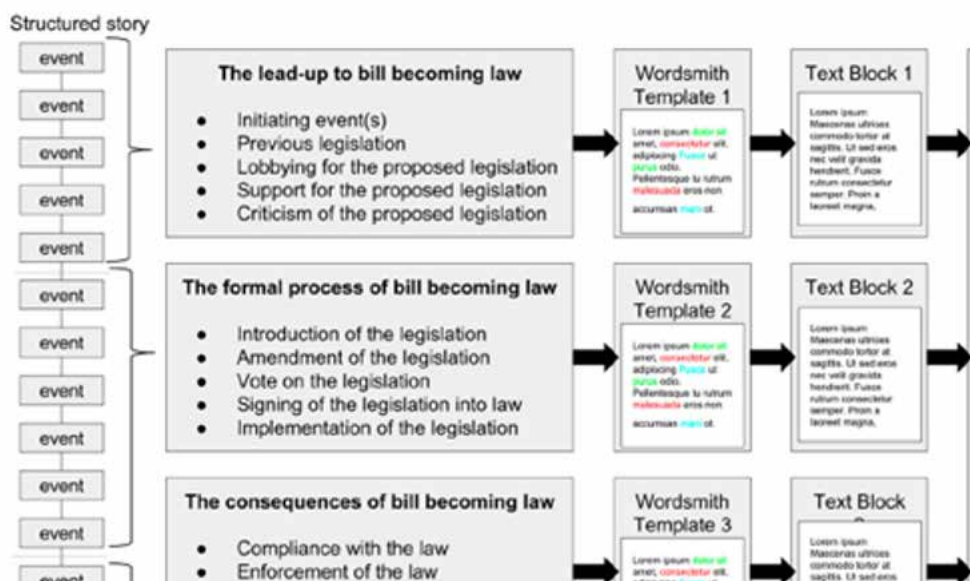


Figura 3. Ejemplo generación de un texto con NLG-estructuras de eventos | Fuente: Caswell y otros, 2018

La consideración sobre los textos periodísticos como un ensamblaje de eventos ha originado un nuevo término, **periodismo estructurado** -en inglés, *structured journalism*-, que se centra en las tecnologías y flujos de trabajo necesarios para recoger información y datos; y, para escribir y publicarlos en contextos periodísticos. Este procesamiento de segmentación también se podría aplicar a la automatización de noticias como videos usando técnicas similares a las utilizadas por la tecnología de producción de video automatizada Wibbitz.com, que actualmente crea videos basados en artículos de texto (Caswell y otros, 2018).

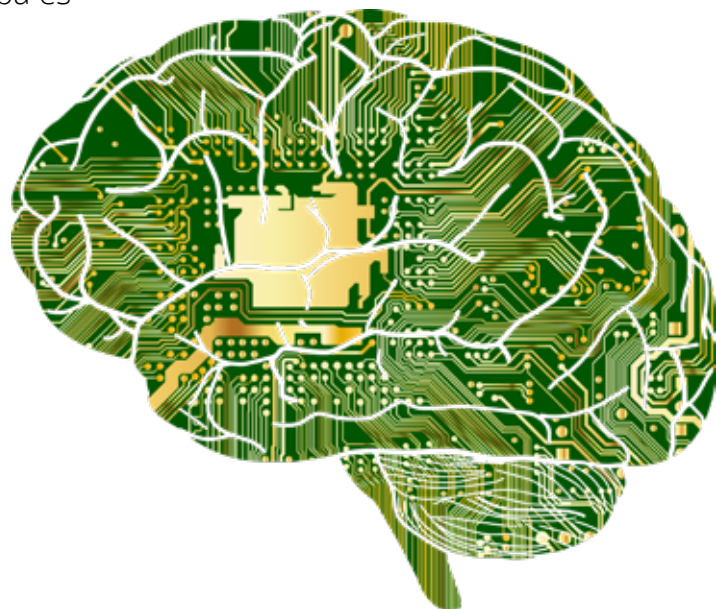
Hay que tener en cuenta que la cantidad de textos periodísticos que se pueden crear con estas técnicas hoy en día es, por ahora, limitada. Pero, se ve claro que, en el futuro, el progreso de la aplicación de la IA al periodismo exigirá crear y formalizar modelos de datos, flujos de trabajo y habilidades editoriales para apoyar la automatización (Caswell y otros, 2018).

Redes neuronales

En este método se usan **modelos de Deep Learning** para generar textos periodísticos. Es decir, se usan redes neuronales para detectar las estructuras de la historia y generar textos periodísticos, o para codificar oraciones y definir modelos que establezcan relaciones entre oraciones o para crear coherencia entre textos periodísticos independientes.

Cuando se usan redes neuronales existe una etapa de aprendizaje o entrenamiento de la red. Esta etapa es importante para que después pueda detectar patrones o tendencias de forma autónoma.

Yao y otros (2019) proponen la generación de textos periodísticos a partir de un título. Así, en la etapa de entrenamiento de la red se utilizan historias existentes de las que se extraerán y generarán automáticamente datos/argumentos de entrenamiento que servirán de referencia (una palabra por cada frase). Finalizado la etapa de entrenamiento, cuando se le proporciona un título a la red, ésta generará datos/argumentos/palabras que corresponderán a la planificación de la historia.



A partir de esos argumentos se escribirá la planificación para obtener la historia. Este proceso correspondería al paradigma estático (ver Figura 4) que presenta Yao y otros (2019) y con el cual ha obtenido mejores resultados. Yao utiliza modelos de generación neuronal para implementar su plataforma y un problema de estos modelos es la repetición de palabras, frases o sentencias en el texto obtenido (ver Figura 5).

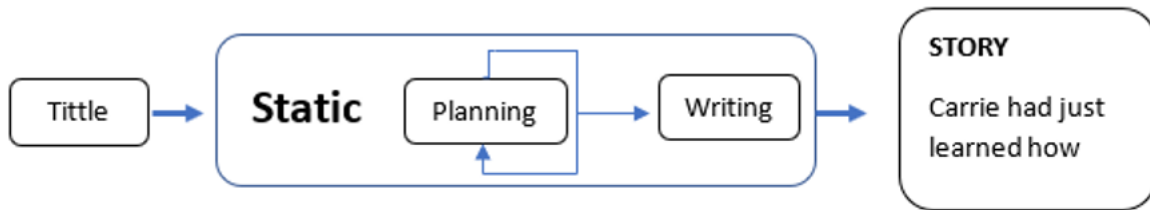


Figura 4. Funcionamiento del paradigma estático | Fuente: Yao y otros, 2019

Title (Given)	The Bike Accident
Storyline (Extracted)	Carrie → bike → sneak → nervous → leg
Story (Human Written)	Carrie had just learned how to ride a bike. She didn't have a <u>bike</u> of her own. Carrie would <u>sneak</u> rides on her sister's bike. She got <u>nervous</u> on a hill and crashed into a wall. The bike frame bent and Carrie got a deep gash on her <u>leg</u> .

Figura 5. Ejemplo texto creado utilizando NLG-Redes Neuronales | Fuente: Yao y otros, 2019





Perspectivas

La automatización de la escritura periodística no es una herramienta de fácil introducción en las redacciones periodísticas. A menudo, suscita entre los profesionales del sector periodístico reacciones negativas, ya que este proceso se ve como una de las principales causas de reducción o destrucción de puestos de trabajo, y como una degradación del trabajo periodístico. Sin embargo, los defensores de la automatización esgrimen, en cambio, que la automatización de la escritura sirve especialmente para superar las tareas repetitivas de los periodistas y facilita que estos puedan dedicar más tiempo a otras tareas más importantes y de mayor valor añadido, tales como el análisis e investigación de los datos y/o de los resultados, entre otras tareas que pueden contribuir a mejorar la calidad del periodismo.

En todo caso, lo cierto es que el impacto de la escritura automatizada en las redacciones periodísticas parece ser

cada vez mayor, y está comportando, sin lugar a duda, serias transformaciones en las tareas profesionales (van Dalen, 2012) y, con toda probabilidad, va a redefinir lo que entendemos por periodismo y por periodismo de calidad.

Las primeras aplicaciones

Hoy en día, es un hecho que tanto compañías líderes como *Associated Press*, *Forbes*, *Reuters* y *The New York Times* –tanto como otras compañías de menor tamaño– están utilizando la IA para generar automáticamente un gran número de textos periodísticos en menor tiempo. Pero, además, las están utilizando para eliminar parte de los errores que se suelen atribuir al factor humano.

Basándose en esta capacidad, muchas empresas periodísticas no dudan en introducir mecanismos de generación automática de textos en sus rutinas de trabajo. Así, la ***Associated Press (AP)***, desde 2014, utilizó la tecnología que les proporcionaba ***Automated Insights*** para generar 4.400 noticias breves –con un máximo 300 palabras– cada trimestre, allá donde los periodistas solo fueron capaces de generar, en el mismo tiempo, solo 300 sin automatización. En 2018, el editor jefe de ***Press Association***, Peter Clifton, anunciaba que se habían escrito 50.000 noticias locales en tres meses con el servicio de noticias de ***RADAR***. Todo ello indica que es muy probable que, en la actualidad, diariamente, muchos miles de textos periodísticos en todo el mundo deban su existencia al concurso de la IA (Lewis y otros, 2019).

...muchas empresas periodísticas no dudan en introducir mecanismos de generación automática de textos en sus rutinas de trabajo...



Ahora bien, no es solo razón de **productividad**, la importancia de la generación automática de textos periodísticos también radica en las ventajas que aporta en relación con la **diversidad**.

De hecho, con el mismo procedimiento, ya descrito, en el que se utilizan bases de datos estructuradas, es posible no solo alterar los datos concretos sino modificar sistemáticamente los enunciados. Por ejemplo, partiendo de la misma frase ya mencionada: **“Mañana en Grenoble, la temperatura media será de 15 grados, y la humedad relativa del aire de 50%”**, podemos imaginar muchas variaciones estilísticas: **“Durante el día de mañana en Grenoble, disfrutaremos de una temperatura media de 15 grados, que se acompañará de una humedad relativa de 50%”**.

Esta vez hemos marcado en negrita los términos que son susceptibles de variación estilística. Basta diseñar, pues, un algoritmo que pueda elegir entre los diferentes términos que se inscriben en un paradigma –una posición sintáctica determinada-. Por ejemplo, se podría elegir entre: 1. “la temperatura media será”; 2. “disfrutaremos de una temperatura media de”; 3. “viviremos una temperatura”; 4. “Los ciudadanos percibirán”; 5. Etcétera. Sencillamente, dispondremos, de esta manera, de una nueva base de datos estructurada conteniendo variaciones estilísticas; y su combinación sistemática con la base de datos de municipio, temperatura y humedad, puede generar –mediante un algoritmo preciso- infinidad de textos periodísticos diversos –o, al menos, de apariencia diversa-. De esta manera, se abre la posibilidad de generar textos periodísticos para ámbitos/ hechos/nichos de audiencia que no se cubrían (Carlson, 2015; Graefe, 2016). He aquí, pues, cómo la diversidad es, también, un valor añadido por la generación automática de textos.



Productividad y diversidad son, pues, las grandes razones por las que el uso de sistemas de generación automática de textos se va a ir incorporando, con facilidad, a las rutinas de ciertas empresas periodísticas. Especialmente aquellas que comercializan cantidades de informaciones a diario -agencias de información general y especializada-, etc. y que se benefician de un modelo de negocio que, al aumentar su productividad, les pueda generar beneficios (Caswell y otros, 2018).

En todo caso, sea cual sea el papel de la escritura automática, la actividad del periodista sigue teniendo una especial relevancia a la hora de seleccionar la información interesante y valiosa, de preguntarse por el sentido de dicha información y por lo que hay detrás de la apariencia de los hechos, de explicar y analizar los datos, y, sobre todo, a la hora de transmitir la información en función del sentimiento y el interés del público. En casi nada de eso, los ordenadores pueden sustituir a las personas (van Dalen, 2012; Carlson, 2015; Haim y otros, 2017).

Imaginemos un caso bien sencillo, a partir de los datos: fresa, fruta, azul, una máquina podría escribir, por ejemplo: “La fresa es una fruta y es azul”. Esta frase desde el punto de vista sintáctico es correcta, lo es también desde el punto de vista semántico porque no contradice ninguna de los sentidos implicados, es decir, es coherente. Sin embargo, no sería aceptable en el contexto que conocemos porque, hasta el presente, las fresas no suelen ser azules. Es ese conocimiento contextual y circunstancial el que a las máquinas les cuesta acertar y, en cambio el periodista aporta un valor singular. En consecuencia, usando la IA en periodismo podemos describir hechos, pero no podemos valorar estos hechos en relación con el contexto ni tampoco garantizar su pertinencia (Carlson, 2015; Haim y otros, 2017). Es posible, pues, que muchos textos generados por máquinas puedan parecer textos producidos por humanos (Lewis y otros, 2019), o que se acerquen cada vez más a ellos (Lewis y otros, 2019). Pero en ningún modo, esto garantizará su pertinencia y su validez contextual.

Por esta razón, a la hora de valorar el trabajo del periodista en relación con el de las máquinas, cuestiones como



credibilidad, veracidad, legitimidad, transparencia y trazabilidad cobran mucha importancia porque la edición automatizada es, en general, oscura y opaca (Diakopoulos, 2015). Lo mismo sucede a la hora de determinar la responsabilidad de una información, no tiene sentido responsabilizar a una máquina de nada de lo que escribe, porque en ningún caso entra en sus parámetros controlar los efectos de lo que escribe o sus consecuencias. Por eso, Shangyuan y otros (2019) defienden la idea de que la tecnología se debería ver y utilizar como un actor más que junto con los periodistas se convierten en “co-creadores”, no en autores. Lo cual no impide, por supuesto, que la tecnología no sea capaz de transformar y alterar los procesos de creación, distribución y consumo de textos periodísticos.

...el peligro de desplazamiento de los periodistas del sistema de producción por parte de la inteligencia artificial no puede dejarse de considerar...

El desplazamiento de los periodistas

En cualquier caso, en un contexto de crisis económica, como ya sucedió en el 2008 y en años posteriores, y como puede suceder tras la pandemia del COVID-19, el peligro de desplazamiento de los periodistas del sistema de producción por parte de la inteligencia artificial no puede dejarse de considerar. Muchas empresas periodísticas al ver recortadas sus fuentes de ingresos se verán tentadas por la posibilidad de sustituir periodistas por máquinas, aumentando, así su productividad, pero dejando de lado el aspecto de la calidad.



Esto, naturalmente, perjudica la calidad periodística. La generación automática de noticias, si no se complementa con la labor de periodistas responsables, capaces de introducir valor al producto elaborado acabará generando contenidos estandarizados que no responderán adecuadamente a las necesidades del contexto. Lo que se perderá si esto sucede es precisamente lo que hace valioso al periodismo de calidad que es la capacidad de análisis y de explicación, y su capacidad de sintonizar con las demandas de los públicos.

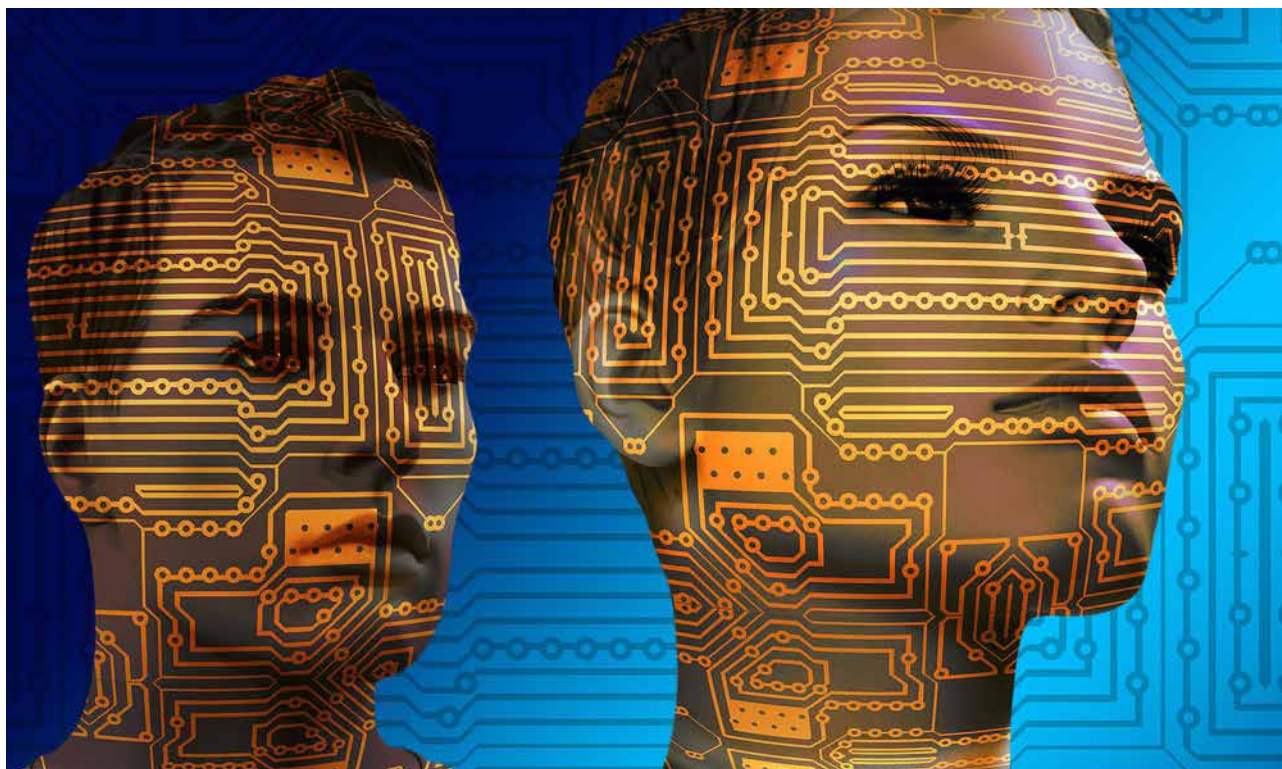
Al mismo tiempo, también resulta evidente que las empresas u organizaciones periodísticas de calidad que no sepan implementar adecuadamente herramientas de generación automática de textos, perderán valor y capacidad de competir en un mercado cada vez más exigente. Lo conveniente, por tanto, es aprovechar todo el potencial de productividad y diversidad de la inteligencia artificial para apoyar, reforzar y promover la labor de los periodistas que tiene que ser, en todo caso, cada vez más sutil, compleja y decisiva. Y que, en última instancia debe asegurar no solo la calidad de la producción, sino su responsabilidad y su valor ético.

La generación automática de textos periodísticos también está abriendo un debate que trata de resituar la generación automática de textos en el lugar que le corresponde y en el contexto de la labor de los periodistas. Son los periodistas los que deben asegurar la transparencia del proceso, y rendición de cuentas, refiriéndose a la responsabilidad de creación de un sistema de IA y de su funcionamiento (Lokot y otros, 2016; Caswell y otros, 2018) y que deberían permitir dar cuenta de su relevancia a la hora de ofrecer un determinado punto de vista y garantizar su objetividad (Gillespie, 2014).

Si **productividad** y **diversidad** eran los argumentos para promover la generación automática de textos periodísticos, la **calidad**, la **responsabilidad** y la **ética** son los valores que impulsan a la defensa de la tarea –insustituible– del periodista.

Entre todos estos valores, con sus tensiones y conflictos, se desarrollará en el inmediato futuro la implementación de la IA en el contexto periodístico.





Herramientas disponibles

¿Cuáles son las principales herramientas de escritura automática que se pueden encontrar en el mercado actualmente? Sin la intención de ser exhaustivos, recopilamos información sobre los servicios comerciales que pueden ser usados en contextos periodísticos.

Para detectar con precisión las aplicaciones y programas que se están utilizando en la actualidad, se creó una base de datos con los principales medios nacionales e internacionales que han implementado el uso de inteligencia artificial en el proceso de elaboración de noticias. Posteriormente se registraron las funciones de cada uno de ellos y se clasificaron en la segunda y tercera fase de la investigación. Se han seleccionado para este informe siete aplicaciones:

- *Heliograf*
- *Cyborg*
- *Wordsmith*
- *Quill*
- *Syllabs*
- *Panels by Getty images and Vizual.ai*
- *Editor*

A continuación, se describen cada una de ellas.



Heliograf



El diario estadounidense *The Washington Post*, a la vanguardia de la tecnología, utiliza desde 2016 **Heliograf**, un robot que aplica la inteligencia artificial para escribir noticias breves sobre política y eventos deportivos. El primer año creó 850 artículos y en 2018 ganó el primer lugar en la categoría *Excelencia en el uso de Bots* de los *Premios Global Biggies*.

El trabajo de **Heliograf** se segmenta en varios aspectos. En el sector financiero, detecta las tendencias y les proporciona a los periodistas una lista de temas con relevancia en la actualidad. En el ámbito deportivo y político es donde más destaca por la capacidad que tiene para recolectar y transformar datos.

Jeremy Gilbert, director de iniciativas estratégicas del diario, asegura en una noticia publicada por *The Washington Post*, que el robot construye un **nuevo modelo para la cobertura local** en cuanto a eventos deportivos se refiere. Gilbert confirma que anteriormente el periódico publicaba solo los juegos con mayor relevancia para la audiencia, mientras que a partir del 2017 **Heliograf** puede cubrir casi de manera instantánea cualquier encuentro del que se dispongan datos.

En política, específicamente en situaciones electorales, el robot es capaz de crear noticias que abarquen temas de interés relacionados con todos los candidatos y puestos a disputar a escala nacional y local. En Estados Unidos, en 2016, se implementó por primera vez para cubrir las elecciones de la

...el primer año creó 850 artículos y en 2018 ganó el primer lugar en la categoría *Excelencia en el uso de Bots* de los *Premios Global Biggies*...

Cámara, del Senado y la gobernación de los 50 estados del país. La intención del diario consistía en publicar informes, análisis y resultados actualizados sobre el evento según la ubicación geográfica de la audiencia.

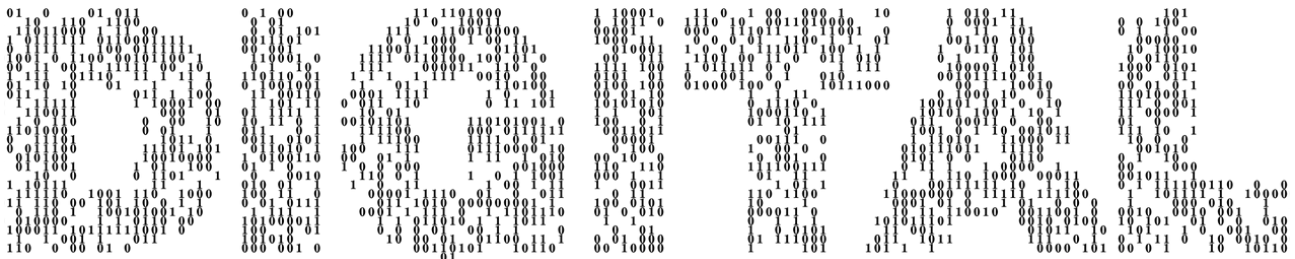


Actualmente, *The Washington Post* desarrolla un laboratorio de periodismo político computacional que servirá de base para implementar novedades en la cobertura de las próximas campañas políticas que se celebrarán en 2020. Se trata de un equipo pequeño, compuesto por unas seis o nueve personas, que trabajarán para implementar nuevas herramientas avanzadas que refuercen los informes y las publicaciones referentes a las elecciones estadounidenses. Entre los pasos de investigación destaca el desglose de datos previo a la ola electoral y la atención de cobertura local de acuerdo con la demanda de la audiencia y de los hechos noticiables.

Cyborg

Bloomberg

Bloomberg es uno de los medios de comunicación que más se ha empeñado en desarrollar herramientas tecnológicas automatizadas con inteligencia artificial que incrementen o faciliten la producción de noticias. *Cyborg* es el nombre del sistema que utilizan para convertir informes financieros en noticias.



A partir de un análisis esquemático, *Cyborg* es capaz de segmentar un documento con datos financieros y crear una noticia con las cifras y los datos más destacados de manera inmediata. De esta forma, cada trimestre, el sistema extrae y convierte los números claves en artículos relevantes. Lo que en un principio era una tarea tediosa para los periodistas, ahora, es un proceso automático y eficaz.



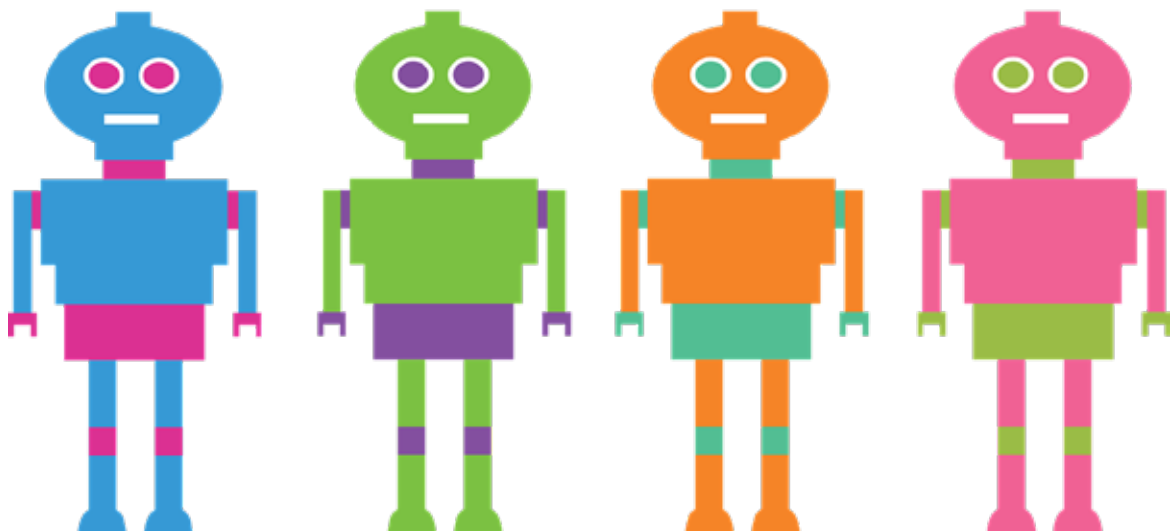
Cyborg no solo le aporta inmediatez al medio de comunicación, sino que además es preciso en los números que publica y coherente en la información que recopila.

Wordsmith



Automated Insights es la empresa estadounidense que creó **Wordsmith**, un software diseñado para automatizar la redacción de artículos y noticias mediante la generación de lenguaje natural (NLG, por sus siglas en inglés).

Wordsmith se encarga de recolectar datos provenientes de diversas fuentes para crear plantillas. Estas plantillas funcionan como formularios prediseñados que se complementan progresivamente con nueva información y dan como resultado la redacción de varios textos. Por lo tanto, el software, de manera inmediata, genera miles de historias basadas en cifras y datos.



Associated Press utiliza esta herramienta para transformar sus bases de datos en noticias e informes publicables. Incluso el sistema fue modificado para lograr una redacción que se asemeje al estilo editorial del medio.



Con **Wordsmith**, AP produce 4.400 historias de ganancias trimestrales o informes financieros, lo que se traduce en un aumento de casi 15 veces sobre los esfuerzos manuales. “Las historias conservan la misma calidad y precisión que los lectores esperan de cualquiera de los artículos escritos por humanos de AP. Además de una nota explicativa al final de la historia, no hay evidencia de que hayan sido escritos por un algoritmo”, aseguran los creadores del software en su sitio web.

Quill



Natural Language Generation (NLG) es uno de los conceptos más relacionados con los softwares y aplicaciones creados para la segunda y tercera fase de la presente investigación. La empresa **Narrative Science** lo toma como eje central de varios productos que desarrolla para automatizar procesos analíticos.





Quill es uno de estos productos. Se trata de una plataforma que crea aplicaciones para recolectar datos y transformarlos en informes. A través de un software de NLG se recrean las funciones de un analista y se redactan documentos con un formato personalizable que se adapta según el usuario y sus necesidades.

En un principio, **Quill** se desarrolló para escribir críticas cinematográficas. En la Universidad Northwestern de Estados Unidos, nació la idea como un experimento en el que el sistema recogía información de sitios como IMDB para crear artículos críticos de películas. Con el tiempo sus funciones se fueron perfeccionando y especializando en lo que es hoy en día: una herramienta de inteligencia artificial dedicada a contar historias partiendo de datos.

Los creadores de **Quill**, ven esta herramienta como una posibilidad no solo para crear noticias con información destacable y actual, sino también para crear informes analíticos y profundos que puedan ser enviados a personas con intereses específicos. Así, el texto sobre la alerta de un fenómeno natural se convierte en un análisis de las consecuencias que este podría generar en una determinada empresa o sector.

De esta manera, en 2018, el PGA Tour, principal circuito norteamericano de golf, encontró en Quill la solución para cubrir el torneo. El objetivo consistió en generar resúmenes del desempeño de los casi 150 jugadores profesionales que competían. Sin el software, por falta de tiempo y capacidad, el equipo de prensa del evento solo publicaba el rendimiento de los golfistas con mejores puntuaciones.

Syllabs



En la constante búsqueda por incrementar el uso de la inteligencia artificial en el periodismo y otros ámbitos, surgen cada vez más empresas y proyectos que revolucionan la tecnología. **Syllabs** es un startup parisino que comenzó su trayecto en 2006 y lanzó su primer programa piloto en 2011.

Hoy en día es un generador de contenido capaz de trabajar en tres idiomas: francés, inglés y español.

En una entrevista con la agencia de noticias Efe, Helena Blancafort, cofundadora de **Syllabs**, explica que estos robots ofrecen agilidad y rapidez porque son capaces de publicar un número de textos notablemente más elevado que los de un equipo humano. Esto, a su vez, incide de manera positiva en el posicionamiento de la web, en la visibilidad y en el alcance de la audiencia interesada.

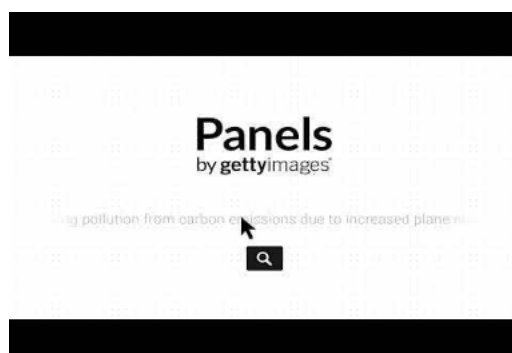
En 2015, el algoritmo de **Syllabs** fue utilizado por el diario **Le Monde** para publicar textos relacionados con los resultados de las elecciones departamentales y regionales en Francia. En total se logró crear contenido para los 34.000 municipios y los 2.000 cantones y cubrir eficazmente las noticias locales. Este logro, imposible para una sala de redacción sin inteligencia artificial, demuestra la magnitud de lo que puede ser capaz una máquina diseñada correctamente.

Syllabs ofrece también un listado de *tags* para repotenciar textos al momento de publicarlos, ampliando así las herramientas de las que dispone su algoritmo. Este complejo sistema para crear contenido no solo es productivo para una empresa comunicacional. En otros ámbitos, como bienes raíces, también se empieza a incluir en la dinámica de trabajo.

El año pasado, la empresa emergente logró recaudar dos millones de dólares de nuevos inversores para ampliar su plantilla laboral e intensificar el desarrollo de nuevos elementos que la podrían categorizar como una empresa líder en inteligencia artificial a escala mundial.



Panels by Getty images and Vizual.ai



La agencia de fotografías **Getty images**, en asociación con la plataforma **Vizual.ai**, integró a su espacio un algoritmo que le recomienda a los editores las imágenes o fotografías más propicias para acompañar una noticia. Con el nombre de **Panels**, esta herramienta de inteligencia artificial funciona mediante filtros personalizados y *deep learning* para lograr una selección acertada.

El algoritmo funciona de manera muy sencilla. En primer lugar, forja una lectura general del texto e identifica las palabras claves. Posteriormente, una vez comprendida la temática, ofrece varias sugerencias de imágenes que puedan destacar el significado del mensaje que se desea transmitir.

La ventaja de **Panels**, y de la mayoría de los sistemas creados hoy en día para optimizar procesos con inteligencia artificial, es que posee una base de datos sólida que le ofrece versatilidad y que por ende incrementa la capacidad de dar una respuesta efectiva.

Vizual.ai explica detalladamente en su sitio web los beneficios de su plataforma: Maximizar la participación del usuario para impulsar una mayor monetización, automatizar las pruebas para garantizar la imagen de mayor rendimiento y simplificar el flujo de trabajo para identificar imágenes atractivas que resuenen con sus clientes.

Sin embargo, debe destacarse que la selección final de la imagen queda a cargo del editor o periodista, por lo que en este caso el algoritmo no asume el trabajo completo del proceso, sino que lo facilita ahorrando esfuerzo y tiempo.



Editor



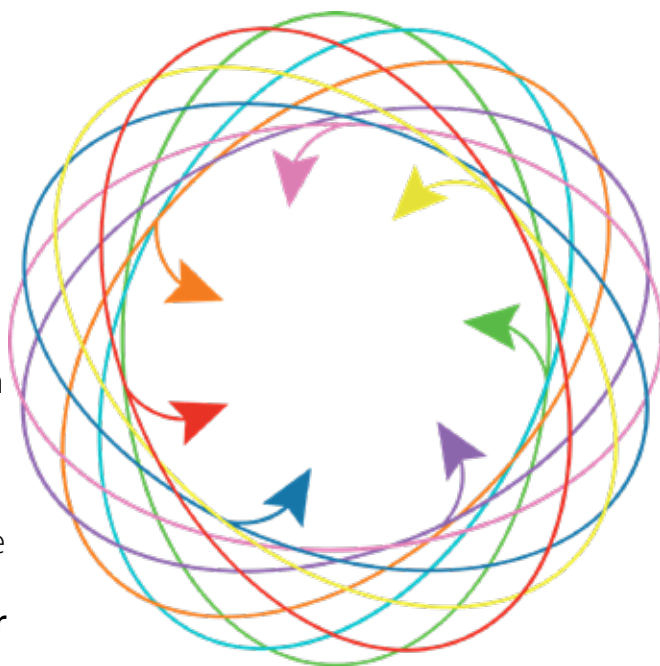
The New York Times se suma a la lista de los medios que, desde hace ya unos años, desarrolla e incentiva la inclusión de la inteligencia artificial en sus salas de redacción. Más allá de generar una noticia o agilizar el proceso de su creación, en 2015 creó un prototipo llamado **Editor** para la revisión automática de textos.

Como parte del proyecto **The New York Times Labs**, grupo de investigación y desarrollo, diseñó un sistema que divide a la pantalla en dos y le permite al usuario o periodista redactar de un lado la noticia y del otro identificar automáticamente frases relevantes, citas o hechos claves para corregir, destacar o eliminar.

“Este prototipo se compone de un editor de texto simple (mostrado a la izquierda), soportado por un conjunto de microservicios en red (visualizados a la derecha). Los microservicios que se muestran aquí son redes neuronales recurrentes que están capacitadas para aplicar etiquetas del New York Times al texto libre”, destaca The New York Times Labs en su sitio web oficial.

El prototipo, conectado a una red, puede a su vez explorar alternativas para fortalecer la escritura, crear nuevos textos y de esta manera aumentar el número de publicaciones. Una herramienta auxiliar que se vuelve cómplice de una labor cada vez más automatizada.

A modo de conclusión sobre las herramientas disponibles en este proceso se podría resumir como la **inteligencia artificial** captura la atención en el mundo del **periodismo**. Un futuro que parecía lejano está en auge y se evidencia en el presente con determinación. Las herramientas tecnológicas no se crean para eliminar puestos de trabajo, como muchos investigadores temían en el pasado, se diseñan para **facilitar y repotenciar procesos** complejos. El reto ahora es usarlas con sabiduría, con **responsabilidad y ética**.





Estudios de caso de escritura automática

En esta ocasión para profundizar en los estudios de caso del proceso analizado se ha escogido la Empresa **Narrativa** dada su referencia en la generación de noticias automatizadas. Para profundizar en su aplicación, se han seleccionado tres medios que ya han integrado este uso en sus rutinas productivas, en concreto: **Agencia EFE, Diario Sport y El Confidencial**.

AGENCIA EFE

<https://www.efe.com>



Agencia EFE, S.A., fundada en Burgos el 3 de enero de 1939 por el entonces Ministro del Interior español Ramón Serrano

Súñer y actualmente presidida por el periodista Fernando Garea, es la principal agencia de noticias en español y la cuarta del mundo, después de Associated Press, Reuters y Agence France-Presse. Se trata de una Sociedad Mercantil Estatal, cuya titularidad corresponde a la Sociedad Estatal de Participaciones Industriales (SEPI), una entidad de Derecho Público adscrita al Ministerio de Hacienda y Administraciones Públicas, con dependencia directa del Ministro.

En 2018, el Presupuesto General del Estado reconoció por primera vez el presupuesto de EFE como un Servicio de Interés Económico General, aumentando la dotación pública de la agencia.

EFE fue la primera agencia española en tener delegaciones en todas las comunidades y ciudades autónomas españolas. En la actualidad cuenta con delegaciones en cada capital de las 17 comunidades autónomas, Ceuta y Melilla, y subdelegaciones en otras ciudades españolas, además de 180 ciudades de 110 países del mundo y cuatro mesas de edición en Madrid, Bogotá, Cairo y Rio de Janeiro. Su misión es ofrecer la visión hispana del mundo en español, portugués, inglés, árabe, catalán y gallego.

EFE cubre todos los ámbitos de la información en los soportes informativos de prensa escrita, radio, televisión e Internet, distribuyendo alrededor de tres millones de noticias al año a más de dos mil medios de comunicación, por medio de una red que cuenta con más de tres mil periodistas de sesenta nacionalidades. En 1995, la agencia recibió el Premio Príncipe de Asturias de Comunicación y Humanidades en reconocimiento a su trabajo, independencia y neutralidad.



EFE y la IA

El equipo de transformación digital de la *agencia EFE*, dirigido por Juan Varela, Director de Estrategia Digital y Negocio, tiene su sede en Nueva York. Esta *task force* tecnológica ya cuenta con un acuerdo con la empresa Narrativa para activar desarrollos de escritura automática, ofrece servicios personalizados de visualizaciones y actualmente está empezando a implementar un programa piloto para la detección de bulos y explorando opciones para la generación automática de videos.

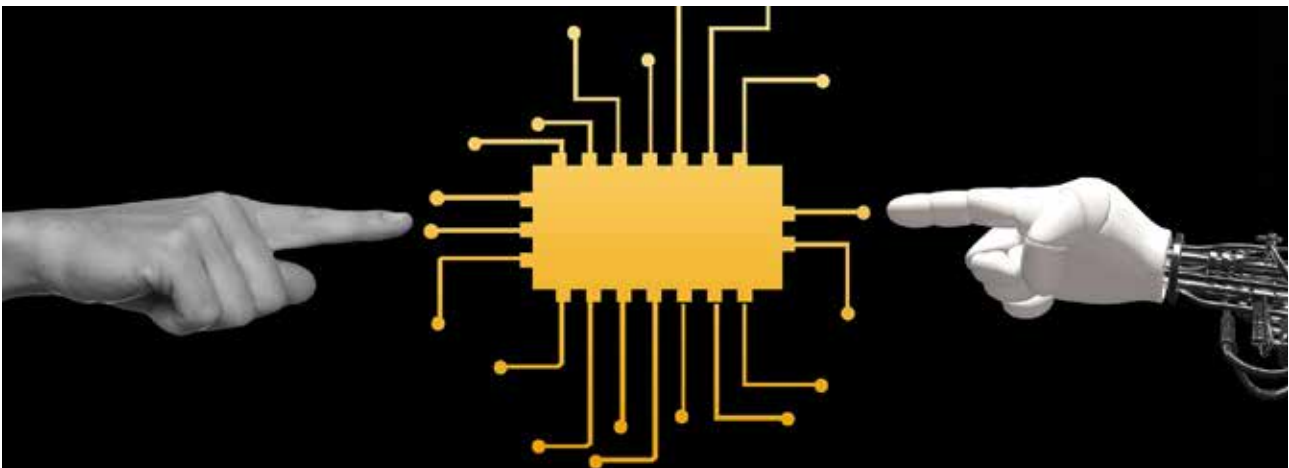
El reto por solucionar

Por un lado, EFE necesita cubrir mucho terreno informativo con un número limitado de recursos humanos y tiempo. Por ello, necesita mejorar los procesos productivos en términos de eficiencia y liberar los periodistas de todos aquellos procesos tediosos que se podrían realizar automáticamente. Asimismo, EFE decide apostar por la escritura automática ya que la mayoría de las agencias ya lo viene haciendo desde hace tiempo. El objetivo final sería llegar a una hiperlocalización de la información.

EFE y Narrativa

Jairo Mejía, periodista con experiencia previa como corresponsal en Tokio, Washington y Seúl y analista de datos para empresas como Idealista o BBVA, representa el enlace entre narrativa y EFE. A diferencia de los anteriores casos, *Narrativa* no solo proporciona un servicio de escritura automática a EFE, sino que las dos empresas tienen un acuerdo abierto que permite a EFE “customizar” el contenido generado por Narrativa, y “escalarlo”, vendiéndolo de forma personalizada a sus clientes.

...el objetivo es llegar a ofrecer una suerte de servicios Premium de longtale, que, cubriendo mucho territorio y temas diversos, permita llegar potencialmente a todos los nichos generando así más impacto...



El objetivo es llegar a ofrecer una suerte de servicios *Premium de longtale*, que, cubriendo mucho territorio y temas diversos, permita llegar potencialmente a todos los nichos generando así más impacto. El modelo de negocio a seguir, según Mejía, es el de *Netflix*, capaz de crear mucho contenido de “nicho”, aparentemente no muy rentable, pero que permite aumentar el impacto cubriendo targets diversos.

Funcionamiento

Como ya hemos visto para los otros casos, **el software Gabriele**, genera contenido a partir de cualquier fuente de datos estructurados (ficheros binarios, datos numéricos, bases de datos, etc.). El software lee los datos fuentes, analiza y extrae la información, a través del análisis de la estructura del lenguaje, y, alimentándose de noticias parecidas, utiliza las informaciones recopiladas para crear narraciones en lenguaje natural, agregando un contexto rico, automáticamente y en tiempo real.

En el caso específico de EFE, *Narrativa* proporciona contenido para deporte, economía, tiempo y pronto empezarán con coberturas electorales. Si bien *Narrativa* proporciona un producto periodístico “llaves en mano” (compuesto por feed rrs, texto e infografía, haciendo la intervención humana innecesaria), EFE tiene un equipo de desarrollo que genera para cada cliente un texto único, ofrece un trato personalizado, que, en las palabras de Mejía, corresponde con la superación del teletipo.

Ejemplo:



EL OLYMPIAKOS PIREUS SE LLEVA LA VICTORIA FRENTE AL ALBA BERLÍN POR 80-99

Con este resultado, los visitantes son duodécimos en la competición

El **Olympiakos Pireus** ganó al **ALBA Berlín** a domicilio por 80-99 en la décima jornada de la Euroliga. Los locales vienen de conseguir la victoria en casa frente al **Kk Crvena**



EL OLYMPIAKOS PIREUS SE LLEVA LA VICTORIA FRENTE AL ALBA BERLÍN POR 80-99

Con este resultado, los visitantes son duodécimos en la competición

El Olympiakos Pireus ganó al ALBA Berlín a domicilio por 80-99 en la décima jornada de la Euroliga. Los locales vienen de conseguir la victoria en casa frente al Kk Crvena Zvezda por 93-80, mientras que los visitantes sufrieron una derrota fuera de casa con el FC Bayern Múnich por 85-82. Con este resultado, el Olympiakos Pireus se queda fuera de los puestos de Play-off con cuatro victorias en 10 partidos jugados. Por su parte, el ALBA Berlín, tras el partido, se queda eliminado de los puestos de Play-off con tres partidos ganados de 10 disputados.

Durante el primer cuarto el liderazgo estuvo en manos del equipo visitante, de hecho, consiguió un parcial en este cuarto de 10-2 y alcanzó una diferencia de 14 puntos (11-25) y acabó con un 17-30. Posteriormente, en el segundo cuarto el equipo visitante se distanció en el marcador y tuvo una diferencia máxima de 22 puntos (27-49) durante el cuarto, el cual concluyó con un resultado parcial de 21-25. Tras esto, los jugadores acumularon un total de 38-55 puntos antes del descanso.

En el transcurso del tercer cuarto el Olympiakos Pireus se distanció de nuevo en el marcador, amplió la diferencia hasta un máximo de 26 puntos (43-69) y terminó con un resultado parcial de 19-21 y 57-76 de total. Finalmente, durante el último cuarto mantuvo su diferencia el Olympiakos Pireus y el cuarto finalizó con un resultado parcial de 23-23. Tras todo esto, los rivales cerraron el marcador del partido con un resultado de 80-99 para los visitantes.

El próximo partido del ALBA Berlín será contra el BC Zalgiris Kaunas en el Mercedes Benz Arena, mientras que en el próximo partido, el Olympiakos Pireus se enfrentará al AX Armani Exchange Milan en el Peace And Friendship Stadium.

Resultados

Jairo Mejía dice estar satisfecho de la experiencia derivada de la estructura automática. En el ámbito comercial ha generado una buena respuesta, permitiendo a la agencia



llegar a clientes que antes no podían ofrecerle servicios tan específicos, por falta de tiempo y de recursos. Los ejemplos concretos que cita son la cobertura de diferentes ligas de béisbol o las predicciones meteorológicas para Córdoba, en Argentina, entre otros casos. Por lo tanto, les ha permitido generar más ingresos. Según Mejía, el periodismo siempre ha tenido mucha información desestructurada, en cambio la IA la tiene la capacidad de **“aunar un montón de información estructurada que se puede utilizar y monetizar”**.

Aceptación en la redacción

Contrariamente a los otros casos analizados, la recepción de esta tecnología entre las redacciones de EFE ha sido rápida y muy positiva. Según Jairo Mejía, **“un periodista de agencia entiende muy rápido lo que se puede quitar de encima”** ya que está acostumbrado a trabajar con ritmos cerrados y a desarrollar tareas tediosas de recopilación de datos. La implementación de la estructura automática les permite, por lo tanto, evitar las tareas más aburridas para dedicar más tiempo a la investigación.

Ventajas y desventajas

La principal ventaja de la escritura automática, según el Project manager, es indudablemente el coste, especialmente el hecho que permita abaratar costes, al mismo tiempo de llegar a audiencias más amplias. La única desventaja es la incapacidad de ser creativa: la IA, de hecho, puede imitar, y replicar, pero no crea algo nuevo. Finalmente, como ya mencionado, este tipo de escritura automática suele alimentarse de datos estructurados. En realidad, sería posible estructurar datos no estructurados, pero el coste superaría los beneficios.

...la principal ventaja de la escritura automática, según el Project manager, es indudablemente el coste, especialmente el hecho que permita abaratar costes, al mismo tiempo de llegar a audiencias más amplias...



Errores y mejoras

Los posibles errores, según el entrevistado, no son más que “fallos técnicos solucionables” y la IA promete acabar para siempre con los mencionados procesos tediosos para liberar al periodista que podrá dedicarse a trabajos más periodísticos (como aspectos de investigación) y menos a la simple recopilación de datos. En términos de mejora, se plantea que la escritura automática vaya relacionada con alertas y que aprenda a sacar datos de las redes sociales, donde hay mucho potencial informativo. Otro aspecto de mejora es que los periodistas empiecen a “entender los datos” a través de formación específica que las empresas deberían ofrecer.



El Futuro

Según Mejía es “imposible que IA y periodismo no vayan de la mano” ya que, especialmente para las agencias que tienen que cubrir mucho territorio con un número limitado de recursos humanos, la única alternativa para sobrevivir es la IA. Tal y como subrayan muchas investigaciones, el futuro parece ser representado por el “Augmented Human”, el ser humano aumentado. En este caso, el entrevistado propone la categoría del “periodista aumentado” capaz, gracias a la ayuda de la IA, de analizar muchos datos y cubrir numerosos territorios a la vez. Por ejemplo, en el caso de que un periodista tenga que escribir una noticia urgente relacionada con la empresa Apple, la IA podría rápidamente generar un histórico de datos financieros o de crecimiento de la empresa, que ayudarían al periodista en la contextualización de la noticia. Sin embargo, el entrevistado subraya que las tareas de análisis y las tareas relacionales (con fuentes, entrevistados, etc.) siempre las tendrán que desarrollar los periodistas humanos ya que “aunque nunca hay que decir nunca jamás, no va a haber un *robocall* que llame y entreviste”.



DIARIO SPORT

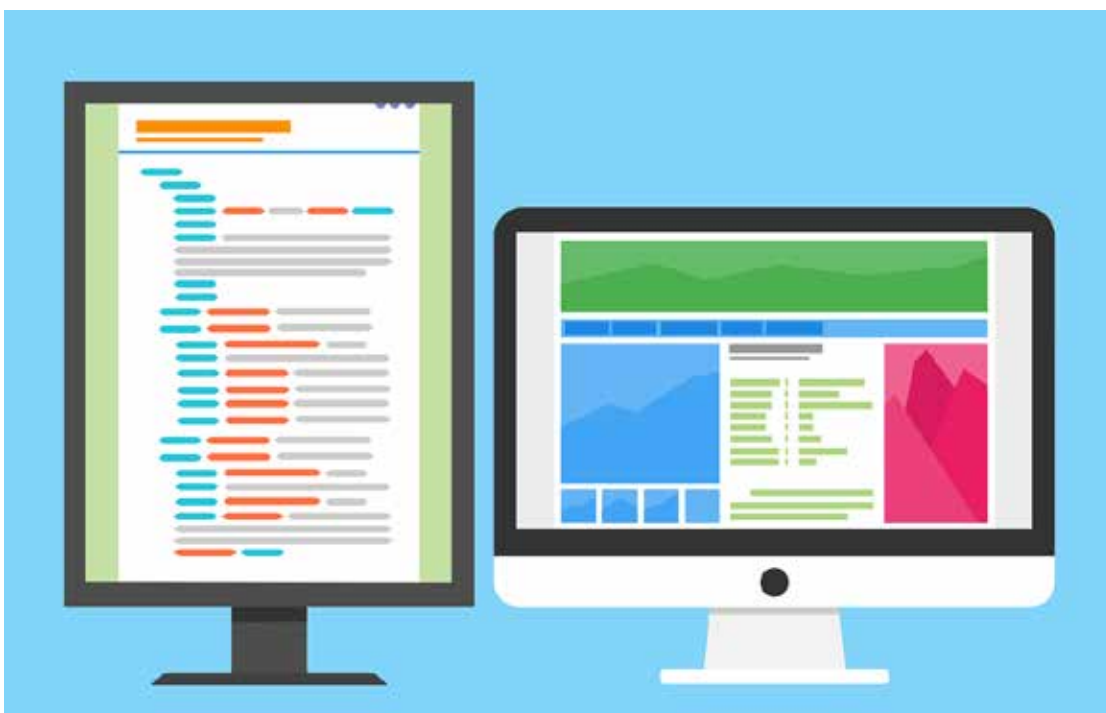
<https://www.sport.es>



Sport es un diario español de prensa deportiva, editado en Barcelona y perteneciente al Grupo Zeta (adquirido por Prensa Ibérica), empresa editora de *El Periódico de Cataluña*. Según los datos del estudio de la Oficina de Justificación de la Difusión (OJD) de julio de 2014, tiene una tirada promedio de 96.981 ejemplares diarios, y una difusión promedio de 61.981 ejemplares diarios. Estos datos lo convierten en el cuarto diario deportivo más vendido en España, por detrás de *Marca* y *As*, editados en Madrid, y de *Mundo Deportivo*, también editado en la capital catalana.

En el año 1997, la empresa decidió sumarse a las nuevas tecnologías, primero de forma modesta y cada vez con más recursos, apostando por la versión digital del diario en internet. Según datos de OJD, la web obtiene más de seis millones de usuarios únicos, más de veinticinco millones de visitas y más de noventa y seis millones de páginas vistas, situándose en primer lugar de la prensa deportiva catalana online.

...en el año 1997, la empresa decidió sumarse a las nuevas tecnologías, primero de forma modesta y cada vez con más recursos, apostando por la versión digital del diario en internet...



El reto por solucionar

El diario *Sport* se dirige principalmente a los aficionados del Barça y, por eso, su difusión tiene un gran alcance en Barcelona y en Cataluña general. La versión online brinda la oportunidad de generar tráfico en todo el territorio nacional. Sin embargo, por su característica de “diario cercano al Barça”, recibe visitas principalmente de Barcelona, resto de Cataluña y de Madrid (por los aficionados madridistas), presentando un déficit de tráfico en el resto de España. Si consideramos que COMSCORE, el medidor de tráfico aprobado por AEDE, tiene panelistas repartidos en toda España, los resultados perjudican los números de este diario. En este sentido, según cuenta Xavier Ortuño, el director adjunto del diario, el desafío es aumentar el tráfico en el resto del territorio español, caracterizado, a menudo, por la ausencia de equipos de Primera División y una demanda de información sobre divisiones menores. Idealmente, por lo tanto, cubrir las ligas menores haría aumentar el tráfico en el resto de España, pero es una tarea muy cara que el diario no podría permitirse.



Narrativa y la aplicación de la IA

En la entrevista personal, Xavier Ortuño, director adjunto diario *Sport*, explica que, en 2018, a través de un director digital Grupo Z, les presentó *Narrativa*, una empresa dedicada a la Generación de Lenguaje Natural (GLN) a partir de datos estructurados y procesables, que les brinda la oportunidad de cubrir las Ligas Regionales a través de contenido generado completamente por un robot.

El Grupo no compra este producto. Sin embargo, *Sport* decide apostar por la inteligencia artificial. *Narrativa* empieza a generar contenido para la cobertura de la Segunda B, luego la Tercera y como los resultados son plenamente satisfactorios, empiezan a generar contenido sobre Ligas extranjeras *on demand*.

A través de una agencia externa, cubre manualmente los principales equipos de las mayores Ligas europeas, pero, por coste, es incapaz de cubrir todos los partidos. En Italia, por ejemplo, se cubren manualmente los partidos de Inter,

...*Sport* decide apostar por la inteligencia artificial...

Milan, Juventus y Roma, dejando al lado los otros equipos. Sin embargo, como explica Ortuño, algunos partidos de equipos no contemplados en este elenco pueden resultar interesantes por razones variadas (presencia de un ex o un futuro jugador del Barça, hechos excepcionales, etc.) y la tecnología les ayuda a cubrir estos vacíos. En un futuro muy próximo, además, se plantea la posibilidad de encargar a *Narrativa* la cobertura de baloncesto y tenis, con la intención de aumentar el volumen de visita a su homepage.

Funcionamiento

Narrativa, a través del software *Gabriele*, genera contenido a partir de cualquier fuente de datos estructurados (ficheros binarios, datos numéricos, bases de datos, etc.) como se ha visto en el anterior caso. El software lee los datos fuentes, los analiza y extrae la información, a través del análisis de la estructura del lenguaje y alimentándose de noticias parecidas. De este modo, utiliza las informaciones recopiladas para crear narraciones en lenguaje natural, agregando un contexto rico, de generación automática y en tiempo real. En otras palabras la empresa proporciona un producto periodístico "llaves en mano", con *feed rrs*s, texto e infografía.



En el caso del fútbol, *Gabriele* se alimenta de los datos de la empresa B-SOCCER, que recopila y digitaliza las fichas técnicas arbitrales o de la federación de prácticamente todas las divisiones (principalmente para su venta en las casas de apuesta). *Gabriele* analiza la ficha técnica y a partir de un entrenamiento a base de noticias similares, agenda, crónica de partidos anteriores y clasificación, produce una crónica a posteriori. Aproximadamente 10 minutos después del final del partido, *Narrativa* entrega un producto acabado en forma de url, una feed rss de contenido bruto que se carga automáticamente en el gestor, que, a su vez, lo publica directamente.

La intervención humana es, por lo tanto, innecesaria: los desarrolladores solo tienen que monitorear. Las crónicas generadas pasan por el filtro de un técnico solo la primera vez y una vez al año para monitoreo.

Ejemplo:



La Nucía consigue el ascenso a Segunda División B después de ganar la final contra el Linares Deportivo por un resultado de 2-1.

Durante la fase anterior del campeonato, las semifinales, La Nucía eliminó a la Arandina, mientras que el Linares Deportivo fue capaz de apearse de la competición al Moralo. Con la victoria de hoy, La Nucía logra el ascenso a Segunda B, donde jugará la próxima temporada, mientras que el Linares Deportivo, por su parte, permanecerá en Tercera División el próximo año.



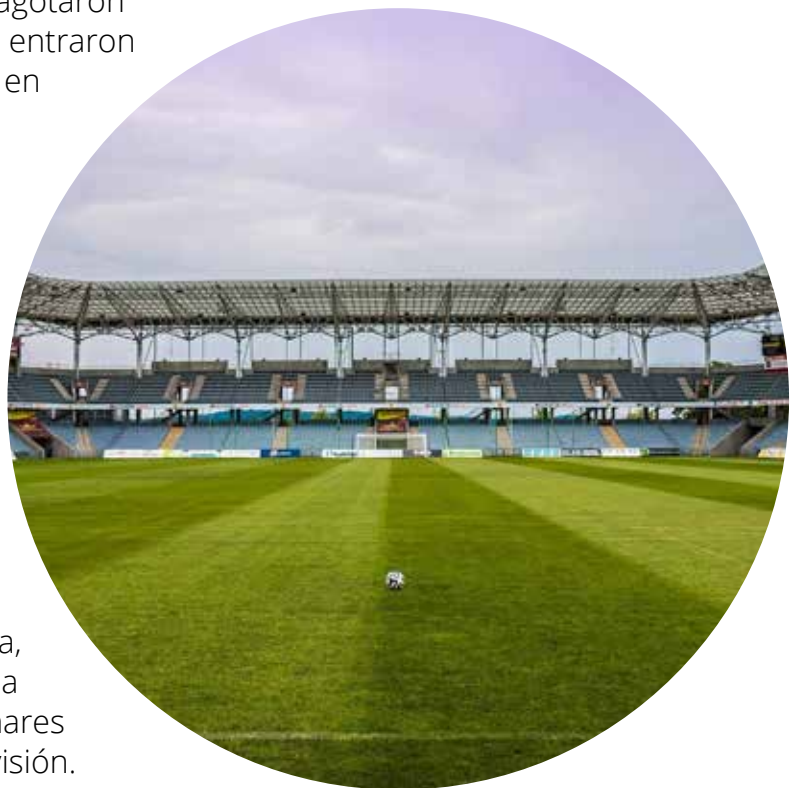
Durante el primer tiempo no hubo goles por parte de ninguno de los jugadores de cada equipo, por lo que el resultado se mantuvo 0-0 durante los 45 primeros minutos.

La segunda mitad del enfrentamiento comenzó de cara para el Linares Deportivo, que estrenó su cuenta goleadora con un tanto de Javi Bolo en el minuto 55. Pero posteriormente el equipo local empató gracias al gol de David Torres a los 71 minutos. Sumó otra vez el conjunto local gracias a un gol de San Julián al borde del final, en el 93. Finalmente, el partido llegó a su fin con un 2-1 en el marcador.

Durante el partido, ambos entrenadores agotaron todos sus cambios. Por parte de La Nucía entraron desde el banquillo Carlos, Mikel y Roserie en sustitución de Jandro, Goku y Javi Salero, mientras que los cambios por parte del Linares Deportivo fueron Pedro Beda, Anaba y Álex Rubio, que entraron por Javi Bolo, Sergio Chinchilla y Pablo Ortiz.

El árbitro sancionó con tarjeta amarilla a tres jugadores. Mostró dos tarjetas amarillas a San Julián y David Torres, de La Nucía y una a Josema, del Linares Deportivo.

Con este resultado, esta final de los *playoff* concluye con la victoria de La Nucía, que se hace con el ascenso a Segunda B la temporada que viene, mientras que el Linares Deportivo jugará de nuevo en Tercera División.



Resultados

El director adjunto se declara plenamente satisfecho ya que han ido subiendo en *ComScore* y han ascendido en el territorio español. En Cantabria, por ejemplo, donde no existen equipos de Primera División, han obtenido resultados remarcables.

Aceptación en la redacción

Por un lado, la IA ha sido muy bien aceptada por parte de los desarrolladores por su facilidad de integración. Por el otro, la mayoría de los periodistas no se han dado cuenta de la integración de la IA ya que cubre un sector de noticias que ellos no tratan. Sin embargo, se percibe un cierto miedo a la

“mecanización” de su trabajo. Aunque la IA no constituye una amenaza, según afirma el Director, “cuando hay un error les gusta destacarlo”.

Ventajas y desventajas

La principal ventaja que destaca el directivo es relativa al coste: si bien no ofrece número exacto comenta que “por el precio de un redactor tengo todas las ligas”. En segundo lugar, la facilidad: tratándose de un producto “llaves en mano”, que no requiere alguna intervención humana, libera tiempo para los profesionales de la redacción. Finalmente, tratándose de un contenido generado por un robot, no comete faltas de ortografía.

Respecto a las desventajas, aunque Ortuño no vea ninguna desventaja particular, subraya que tratándose de “fichas técnicas narradas” es impensable una sección solo de contenido generado por robot, ya que “falta que la noticia sea del Sport”. Esto es: que esté escrita con “intención y divertimento”. Según los desarrolladores sería posible enseñar a la máquina a “escribir con bufanda puesta”, en el caso del Sport, por ejemplo, entrenándola a mirar desde la perspectiva del Barça, pero sería demasiado caro. Por lo tanto, la apuesta parece ser una mezcla de noticias generadas por el robot con noticias generada por periodistas, caracterizadas por la intencionalidad y la emoción.



Errores y mejoras

La IA no es exenta de errores. Los errores pueden ocurrir cuando “pasan cosas que superan a los robots”, como, por ejemplo, hechos imprevistos: partidos que se aplazan o cancelan por lluvia o mal tiempo, etc. La temporada pasada, por ejemplo, el Reus fue retirado de la competición por impago. Sin embargo, el robot seguía haciendo las previas. Los márgenes de mejoras residen en la diferenciación del contenido, ya que todas las noticias generada por el robot parecen iguales.

No obstante, Ortuño admite que “la máquina está

aprendiendo" ya que si al principio, de hecho, todas las noticias eran iguales, ahora van mejorando.

El futuro

Según Ortuño, la IA se volverá un instrumento muy útil para la generación de contenido de todos aquellos sectores muy "numéricos", caracterizados por datos como datos electorales, bolsa, etc.

Para el periodismo, en particular, será fundamentales para aquellos contenidos geográficamente replicables y serializables. En el caso específico del periodismo deportivo, la IA será una aliada fundamental para cubrir todos los deportes muy numéricos, como por ejemplo el ajedrez o el golf.

Sin embargo, no puede funcionar para deportes como el boxeo donde el factor humano es fundamental. Además, apuesta por una fórmula híbrida en la cual la IA proporciona al periodista el "esqueleto" de las noticias (datos básicos, infografía, etc.) para que un humano lo complete de forma que se pueda adaptar según las diferentes necesidades.

EL CONFIDENCIAL

<https://www.elconfidencial.com>

El Confidencial

El Confidencial es el tercer medio generalista más leído a diario en España, por debajo de *El Mundo* y *El País*. Fundado en 2001 por José Antonio Sánchez García -que en la actualidad cuenta con el 40 % de las acciones- y Jesús Cacho y Antonio Casado Alonso, es una publicación de Titania Compañía Editorial SL y sus empresas subsidiarias Vanitatis SL y Titania Eventos SL.

El diario, originalmente fue pensado como un medio especializado en noticias de carácter económico, destinado a un nicho muy pequeño de lectores relacionados con la banca y el mundo empresarial, con el tiempo fue creciendo e



incorporando diferentes secciones como Cotizalia (economía y mercados), Mundo, Cultura, Teknautas (tecnología y ciencia), Deportes, Alma Corazón y Vida (sociedad y bienestar), Vanitatis (crónica social) y Gentleman (estilo de vida masculino), orientándose hacia un público más general, profesional, de mediana edad y con una orientación política liberal. Su eslogan es “El diario de los lectores influyentes”, que guarda relación con el carácter inicial del medio.

En la actualidad el diario cuenta con unos 200 trabajadores distribuidos entre su redacción principal, en Pozuelo de Alarcón y delegaciones de Barcelona, Valencia, Sevilla, Málaga, Galicia y País Vasco. El diario ofrece todos sus contenidos de forma gratuita en Internet, sin muro de pago o socios. Su financiación procede, especialmente, de la publicidad, eventos y contenido de marca (*branded content*).

El Confidencial y la IA: El Confidencial Lab

Alejandro Laso, periodista e informático, llega al *Confidencial* en 2010, y dos años más tarde, bajo el impulso del consejero delegado, funda Elconfidencial.LAB, el laboratorio de innovación, estrategia de producto y nuevas narrativas de El Confidencial, del cual es responsable. La creación del laboratorio se debe, en sus propias palabras, a la intención de “convertir al medio en una empresa de tecnología, sin perder su esencia”.

En la actualidad el diario cuenta con unos 200 trabajadores distribuidos entre su redacción principal, en Pozuelo de Alarcón y delegaciones de Barcelona, Valencia, Sevilla, Málaga, Galicia y País Vasco.



Alejandro añade que el periódico había pasado por dos etapas: su creación, y la elaboración de un producto informativo diferente, de muchísima calidad y rigor, y la generación de nuevas secciones para completar la oferta informativa y atraer a nuevos lectores. Gracias a haber seguido esos pasos y en ese orden, según Laso, El Confidencial se convirtió en un producto de éxito, con una audiencia sólida con un 70% de 'parroquianos'.

La creación de ElConfidencial.LAB representaría la tercera etapa del periódico, en la que se apuesta decididamente por el "salto digital", la innovación y la tecnología. El laboratorio, compuesto por 30 personas, está dividido en cuatro equipos de trabajo, cada uno de ellos orientados a diferentes productos: CMS, movilidad, desarrollo de audiencias y nuevas narrativas (gráficos, vídeo y periodismo de datos). Cada equipo, que trabaja de forma independiente, está compuesto por un jefe de producto, un diseñador y un equipo técnico. Involucrar al departamento técnico en la estrategia de la compañía, según el entrevistado, es la mayor innovación. Cada proyecto pasa por varias fases: ideación, análisis, creación de prototipos, diseño, refinamiento, desarrollo, despliegue y de análisis de resultados. Ese cóctel, en sus propias palabras, es el que hace que "el 70% del tráfico que recibimos cada día sea directo".

...el Confidencial se acerca a la escritura automática con la intención de "dejar la materia gris para un periodismo más cualitativo"...

El reto por solucionar

Según Alejandro Laso, *El Confidencial* se acerca a la escritura automática con la intención de "dejar la materia gris para un periodismo más cualitativo". A través de la escritura automática, de hecho, es posible liberar tiempo de los periodistas para que dejen de ejecutar las funciones que puede realizar una máquina y dedicarse a tareas más cualitativas, como, por ejemplo, la investigación. Además, el diario tiene la necesidad de "capitalizar" la información, hasta llegar a la plena personalización de las noticias, y la escritura automática permite ahorrar tiempo y recursos.

Narrativa y la aplicación de la IA para escritura automática

ElConfidencial.LAB apuesta por la escritura automática de todas las informaciones basadas en datos estructurados, a través de la *partnership* con la empresa Narrativa, Concretamente confía en Narrativa



para deportes, informaciones de servicio (loterías, tiempo, programación televisiva, etc.) y, desde las últimas elecciones municipales, para la cobertura electoral. En las citadas elecciones, de hecho, el algoritmo de Narrativa cubrió autónomamente los resultados de todos los municipios españoles con más de 5.000 habitantes en tiempo real. Para la implementación de este proyecto el laboratorio creó un “oasis” con periodistas completamente dedicados a entrenar los algoritmos.

...el software lee los datos fuentes, analiza y extrae la información...

Funcionamiento

Narrativa, a través del *software Gabriele*, genera contenido a partir de cualquier fuente de datos estructurados (ficheros binarios, datos numéricos, bases de datos, etc.). El software lee los datos fuentes, analiza y extrae la información. A través del análisis de la estructura del lenguaje y alimentándose de noticias parecidas, utiliza las informaciones recopiladas para crear narraciones en lenguaje natural, agregando contexto, automáticamente y en tiempo real. Como se mencionó anteriormente, al principio todo el contenido de *Narrativa* pasaba por la supervisión de un periodista. No obstante, una vez entrenado, el algoritmo tiene acceso al gestor de contenido a través de una “puerta” y puede subir directamente el contenido. En otras palabras, *Narrativa* proporciona un producto periodístico “llaves en mano”, con *feed rrs*, texto e infografía. El algoritmo se alimenta de fuentes de datos públicas: en el caso de las elecciones, por ejemplo, se alimenta de datos SITEL –entre otros– y es capaz de recopilar reconocer autónomamente los escudos cada Comunidad Autónoma. La intervención humana es, por lo tanto, innecesaria.

Ejemplo:



27/05/2019 14:00 - ACTUALIZADO: 30/05/2019 15:23

Con un 100% de los votos escrutados, el PSOE triunfaría en las elecciones municipales de Getafe con 11 concejales, lo que supondría tres representantes más que en los pasados comicios. La formación socialista habría logrado un 35,14% de los votos, mientras que en la segunda plaza aparecería el PP, que se haría con cinco ediles, cuatro menos que en las anteriores votaciones en esta localidad; seguido de Cs, con cinco. El cuarto lugar lo ocuparía Podemos - Ahora Getafe, que se habría hecho con tres, seguido de Vox —6,49%— que consigue dos y de MMCCG —5,54%— con uno. Tal resultado dejaría sin representación a IGE, opción elegida por el 3,89%; y a Izquierda Unida - Madrid en Pie, con un 2,64%.



Un 68,33% de electores ha acudido a las urnas en el territorio, empeorando así el dato registrado durante la pasada cita electoral, cuando registró el 69,97%. Pese a los niveles generalmente inferiores, en las 35 provincias que suelen celebrar simultáneamente elecciones locales y autonómicas, la participación en ambas ha sido históricamente casi idéntica, a partir de 2011 ha sido algo más alta en las municipales e, incluso, en no pocas circunscripciones el nivel participativo de locales y autonómicas se ha equiparado con el de las generales.



Resultados en la provincia

En el conjunto del país, el PP aparecería como gran triunfador al aventajar a sus rivales en 2563 localidades. La formación popular encabezaría 47 municipalidades de Madrid, lo que le sitúa por encima del PSOE, que sería la formación con más votos en 25 plazas; AINS, en uno; y ATySch, en uno. Estos son los partidos con mayor representación en la provincia:

PP: 762 concejales (796840 votos).

PSOE: 653 concejales (678914 votos).

Cs: 257 concejales (544129 votos).

Vox: 140 concejales (248628 votos).

Unidas Podemos: 64 concejales (82902 votos).

En cualquier caso, estos datos no son definitivos y habrá que esperar al final del escrutinio, que en 2015 se produjo a partir de las 23:30 horas en el conjunto del país, pero que en esta ocasión podría llegar incluso más tarde teniendo en cuenta que los colegios electorales de Italia cierran a las 23:00 horas y, hasta ese momento, no podrá hacerse público el resultado de los comicios europeos, que precede a los locales y a los autonómicos, por este orden. El recuento definitivo lo llevan a cabo las juntas electorales de zona por orden alfabético de municipios y no comienza hasta tres días después, de forma que los resultados oficiales se publican 40 días más tarde en el Boletín Oficial del Estado (BOE).



Resultados

Laso se declara plenamente satisfecho de la experiencia, que define “un inicio esperanzador”. En particular destaca que, especialmente en el caso de la cobertura electoral de las Municipales, el mismo hecho de poder llegar a analizar poblaciones tan pequeñas, ha generado un rápido aumento de volumen de público. Asimismo, subraya la ayuda para el periodismo local y la capacidad de la máquina de ser atemporal, es decir de trabajar noche y día sin descanso.



Aceptación en la redacción

Alejandro Laso destaca que los primeros periodistas involucrados, los de la sección de Deportes, se lo tomaron casi con humor, “como un juego”. No obstante, al constatar que efectivamente funcionaba, en principio, se asustaron pensando que estaban “alimentando un monstruo” que podía quitarles el trabajo. Sin embargo, finalmente entendieron sus funcionalidades, hasta el punto de poner un nombre propio a la IA, Ana Futbol, convirtiéndola en “una más del equipo”. En cambio, en las otras secciones la propuesta de IA genera desconfianza, sobre todo en el caso de la cobertura electoral. Sin embargo, a medida que los periodistas van descubriendo sus ventajas, se van acostumbrando.

Ventajas y desventajas

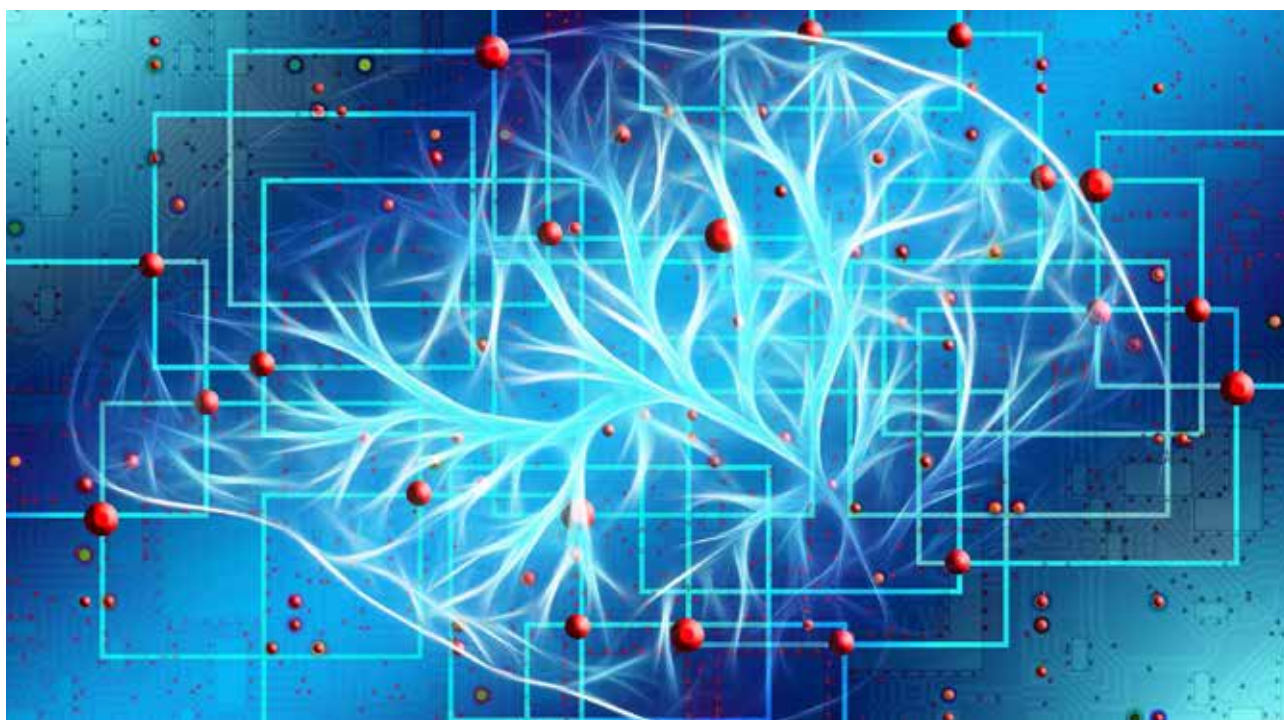
Además del mencionado rápido incremento de volumen de público, Laso insiste en las potencialidades de la escritura automática para el periodismo local, que tiene limitaciones de tiempo y de recursos. Igualmente recalca en las potencialidades de personalización de la información. En el caso del fútbol, por ejemplo, insinúa que no sería difícil entrenar a la máquina para que de la misma noticia dé una versión pro Messi, una pro Cristiano Ronaldo, etc., para poder eficazmente satisfacer todos los targets. Asimismo, en términos de desarrollo, considera que la IA podría ser de gran ayuda para los periodistas en términos de, por ejemplo, algoritmos capaces de monitorear los currículums de todos los políticos, para ayudar en casos de datos falseados, o que analicen datos sobre “puertas giratorias”, así como la creación de bots internos que agilicen los procesos productivos.

En cuanto a las desventajas, recalca que los algoritmos de escritura automática tienen dos limitaciones fundamentales; por un lado, se limitan a datos estructurados, por lo tanto, no funcionan con datos más cualitativos o desestructurados. Y, por el otro, siempre necesitan de la intervención de un humano, en este caso un periodista, para entrenarlos.

Errores y mejoras

Al principio la IA generaba “textos que no estaban a la altura” (hechos de contenido “robótico y poco publicable”) y la máquina necesitaba de la constante intervención de un periodista “entrenador” para dar forma al texto. Sin embargo, Laso admite que “la máquina está aprendiendo” y que va mejorando. Por el otro, los periodistas mismos, todavía no tan involucrados con la tecnología, cometían errores entrenando la IA. En relación con los márgenes de mejora, Laso sugiere que la máquina pueda añadir al paquete de servicios ofertados más contenido audiovisual, de voz para dirigirse a una mejor segmentación y personalización, servicios de geolocalización “embedded”.

...la máquina necesitaba de la constante intervención de un periodista “entrenador” para dar forma al texto...



El futuro

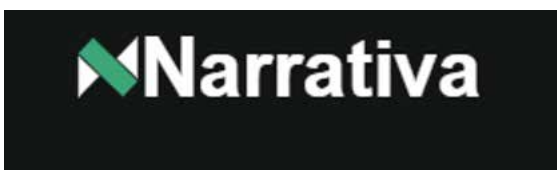
Laso se niega a hablar de IA en futuro, ya que considera que es la realidad que ya está aquí, por lo tanto, si “tú no innovas, lo hará otro y si te relajas enseguida te quedas fuera”. En concreto, según el responsable de El Confidencial Lab, con la llegada de *Facebook Instant Articles*, *Google Play Kiosko*, *Apple*

News y *LinkedIn Pulse*, el lector va a tener a su disposición una oferta informativa enorme, con la cual las aplicaciones puramente informativas de los medios de comunicación no podrán competir. La única oportunidad radica, pues, en crear en productos innovadores que el usuario no pueda encontrar en otro sitio. El objetivo, por lo tanto, es ofrecer una experiencia de navegación personalizada, dependiendo del dispositivo que el usuario use, de su localización y de sus intereses. En concreto se trata de que la IA pueda hacer predicciones generando “ventanas de uso” con sugerencias de temas que, unidas a la geolocalización, permita llegar mejor a los usuarios. Un ejemplo podría ser un sistema de alertas de noticias basado en el interés del usuario que detecte, por ejemplo, que el mismo está viajando en metro y que, por lo tanto, está disponible para recibir y leer contenido.

En conclusión, según su opinión los datos son tan importantes para el periodismo que en breve ya no se hablará de periodismo de datos, sino simplemente de “periodismo”. Por lo tanto, los periódicos que hoy en día no cuenten con periodistas especializados en datos, simplemente se verán obligados a cambiar de idea. Sin embargo, añade que no es fácil encontrar este tipo de perfiles periodísticos en el mercado, ya que hay una laguna enorme entre lo que enseñan en la carrera de periodismo y lo que actualmente piden las empresas periodísticas.

NARRATIVA

www.narrativa.com



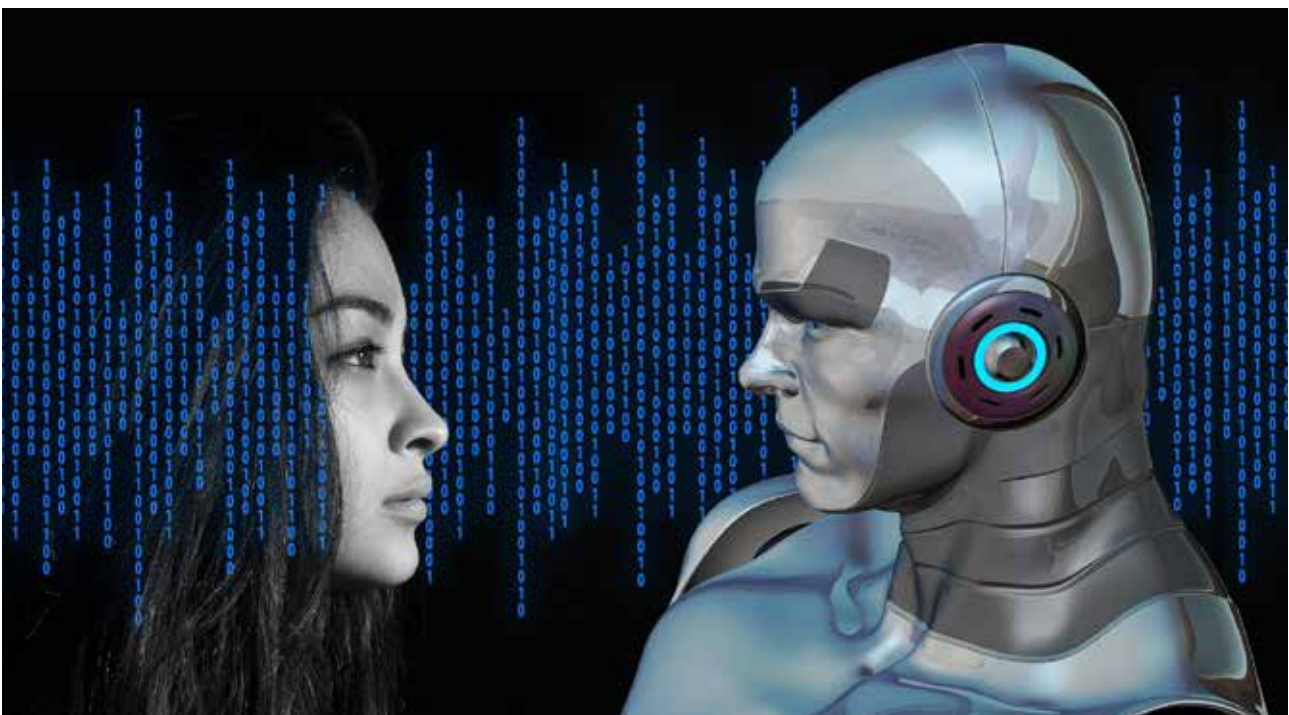
Agencias internacionales como *Reuters*, *Associated Press* o *Bloomberg*, desde hace tiempo desarrollan contenidos basados en datos estructurados e inteligencia artificial. Entre sus planes estratégicos, se encuentra el incremento de este tipo de contenidos. *El Washington Post* ya acudió a la robótica para realizar parte de su cubrimiento informativo de las pasadas Olimpiadas. El primer intento comercial de convertir datos en historias de forma automática fue un proyecto de una empresa clave en el sector, *Narrative Science*, cuyo algoritmo “Quill”, creado en 2010, está programado para aprender el

lenguaje del ámbito que se le asigne. La empresa recién recibió una inyección de capital proveniente de la compañía Q-Tel, un fondo de inversión en inteligencia, para potenciar su cometido. Otros competidores incluyen a la estadounidense *Automated Insights*, cuyo software *Wordsmith* alimenta la generación de contenidos automáticos de *Associated Press*, *Arria*, *AX Semantics* e *Yseop*, entre otros. En Europa RADAR (Reporters And Data And Robots) ha sido dotado de una de sus mayores ayudas hasta el momento, con algo más de 700.000 euros, para ayudar a la agencia *Press Association (PA)* que suministra información al Reino Unido e Irlanda a crear hasta 30.000 noticias locales al mes.

Creada en 2015, *Narrativa* es la primera *start-up* española (con sede en Madrid y Abu Dhabi) dedicada a la creación automática de contenidos a través de la IA (inteligencia artificial).

Los fundadores

Los fundadores de *Narrativa* se conocieron en el departamento de ciencias Informáticas de la Universidad de Alcalá de Henares, donde ambos trabajaban. En particular, David Llorente, licenciado en Física e Ingeniería Informática y actualmente Doctorando en Inteligencia Artificial y generación de lenguaje natural, ha fundado 4 startups. Antes de su carrera como emprendedor, Llorente había trabajado en el desarrollo de negocios para compañías líderes en el sector del juego online en Europa, Estados Unidos y Latinoamérica.



La historia

David Llorente trabajaba en una empresa de juego online que tenía problemas de generación de contenido. La generación de contenidos para los jugadores, de hecho, ocupaba mucho tiempo que afectaba a la labor de los desarrolladores. De aquí la idea de Narrativa, que nace en 2015, con la misión de utilizar los datos disponibles para generar texto a través de un software creado ad hoc y llamado *Gabriele*.

La tecnología: Gabriele

Gabriele es un software para la Generación de Lenguaje Natural (GLN) a partir de datos estructurados y procesables. Gabriele puede trabajar con una gran variedad de sectores de negocio, ya que puede generar lenguaje a partir de cualquier fuente de datos estructurados (ficheros binarios, datos numéricos, bases de datos, etc.). El software lee los datos fuentes, analiza y extrae la información. A través del análisis de la estructura del lenguaje y alimentándose de noticias parecidas utiliza las informaciones recopiladas para crear narraciones en lenguaje natural, agregando un contexto rico para un significado más profundo y palabras de estilo para el idioma y el tono locales, automáticamente y en tiempo real. Por ejemplo, si se alimenta de noticias sobre el desempleo, a través de sus algoritmos extrae las frases, las clasifica por temas por similitud y aprende como escribir sobre ese tema.

Demo: <http://demo.narrativa.com/>

Validación y fase de entrenamiento

Este proceso se puede definir como semiautomático ya que, como subraya Llorente, “la IA no es fiable al cien por cien, y siempre hay un margen de error”. Por ello, siempre hay un elemento humano. La “maquina” se entrena con bases de datos y cuando produce los primeros outputs unos periodistas validan y corrigen hasta que el sistema sea capaz de operar autónomamente.

Sectores

Gabriele puede trabajar con cualquier sector que se alimente de datos estructurados: los deportes, las finanzas, la meteorología, los datos estadísticos, los sondeos o resultados políticos, entre otros. Los principales sectores de aplicación de este sistema son:



a) Servicios Financieros

Debido al gran volumen de datos, bancos y aseguradoras corren el riesgo de no identificar e interpretar la información apropiada a tiempo: *Narrativa*, puede convertir los datos en informes y en comunicaciones personalizadas para audiencias concretas. Es decir, en lenguaje natural de fácil comprensión.

b) E-Commerce

El sector del e-commerce se caracteriza por un alto coste en la generación y en la traducción de descripciones de productos, por los frecuentes errores, por la falta de consistencia en las descripciones de productos y por el riesgo de posicionamiento SEO pobre al usar descripciones provistas por los fabricantes. *Gabriel* puede generar automáticamente descripciones de producto únicas basadas en las especificaciones y en varios idiomas.

c) Sanitario

El sector sanitario tiene como reto el de “traducir” el lenguaje especializado en uno que los pacientes puedan entender. A partir de grandes volúmenes de datos y fuentes que han de ser analizados y combinados de forma eficiente, *Narrativa* puede generar informes y otras formas de comunicación basadas en la misma fuente de información, pero con distintos lenguajes y estilos para diferentes audiencias: personal médico, pacientes, aseguradoras, etc.

d) Inmobiliario

El sector inmobiliario incurre en costes importantes para generar descripciones de inmuebles que pueden contener errores o no estar redactadas de forma adecuada. La IA puede generar descripciones de inmuebles con alta calidad a partir de los datos de éstos.

e) Telecomunicaciones

Los servicios de telecomunicaciones mueven enormes cantidades de datos, interactuando de forma muy frecuente con grandes cantidades de clientes con problemas de muy diversa índole (altas, activaciones, facturación, planes, redes, etc.). *Narrativa* es capaz de generar comunicaciones personalizadas para cada cliente teniendo en cuenta la historia completa del servicio.



Las soluciones de Narrativa analizan automáticamente grandes cantidades de datos para identificar los problemas relevantes y comunicarlos en lenguaje natural de fácil comprensión.

Narrativa para los medios de comunicación

Como se mencionó anteriormente, *Gabriele* puede generar contenidos periodísticos para todos aquellos temas que se basen en datos estructurados como deportes, finanzas, meteorología, estadísticas, sondeos y resultados políticos. Dicho de otra forma, *Gabriele* puede sustituirse al periodista en todas las tareas repetitivas y automatizables. De este modo, un medio puede generar más contenido a un coste menor, superando los obstáculos de la baja productividad del equipo editorial y del alto coste en generar contenido único para cada audiencia. La mayor generación de contenido, a un coste menor, permite llegar a audiencias nicho, mejorando el posicionamiento SEO. Por el otro lado, el equipo editorial, liberado de las tareas mencionadas, puede dedicar su tiempo a contenido más complejo.

...*Gabriele* puede sustituirse al periodista en todas las tareas repetitivas y automatizables...



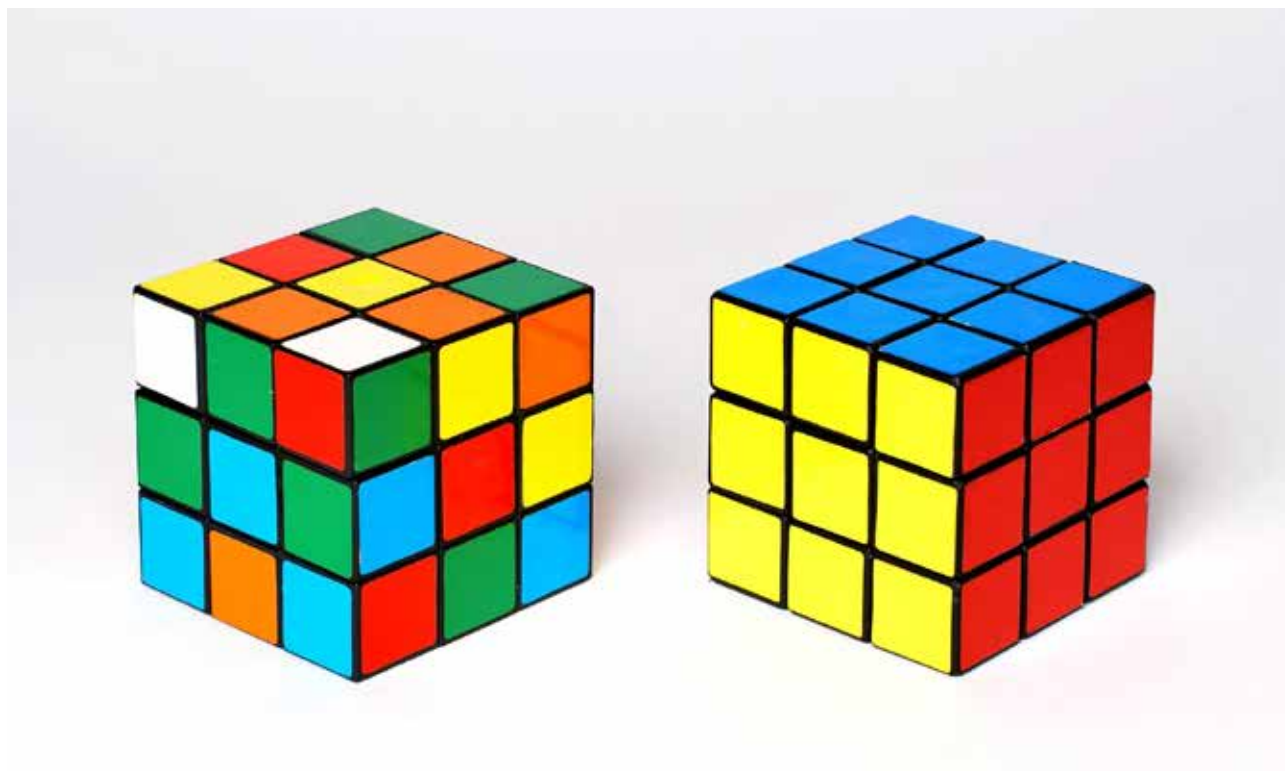
Clientes

Entre los clientes de *Narrativa* encontramos *MSN*, *el diario Sport* y *El Periódico*. Además, en julio 2019, la Agencia EFE y *Narrativa* firmaron un acuerdo para lanzar una ambiciosa oferta de contenidos automáticos. EFE se convertirá así en distribuidor y comercializador exclusivo de los productos de *Narrativa*, con la que desarrollará sus productos actuales y explorará

nuevas oportunidades. Con esta alianza, EFE incorporará los contenidos generados por inteligencia artificial en su oferta a los clientes y en el desarrollo de sus propios productos, como ya hacen las grandes agencias internacionales. EFE trabajará con *Narrativa* en el desarrollo de contenidos destinados a la audiencia y los clientes en áreas como los deportes, las finanzas, la meteorología, los datos estadísticos, los sondeos o resultados políticos. Además, aprovechará la inteligencia artificial para procesar datos, tanto de fuentes estructuradas como de internet y otras plataformas, para ofrecer a los periodistas y editores material seleccionado, ordenado y jerarquizado sobre el que trabajar para construir contenidos de alto valor añadido.

Retos

En la entrevista realizada a David Llorente, el CEO subraya que el periodismo está en crisis y la profesión “está cambiando”. Sin embargo, todavía persisten ciertas resistencias, sobre todo entre los periodistas de edades más avanzadas que no están familiarizados con las TIC y que temen perder su trabajo. Con relación a ello, Llorente defiende que “nadie va a perder su trabajo”, ya que la maquina hará el trabajo del periodista más ágil y menos repetitivo.



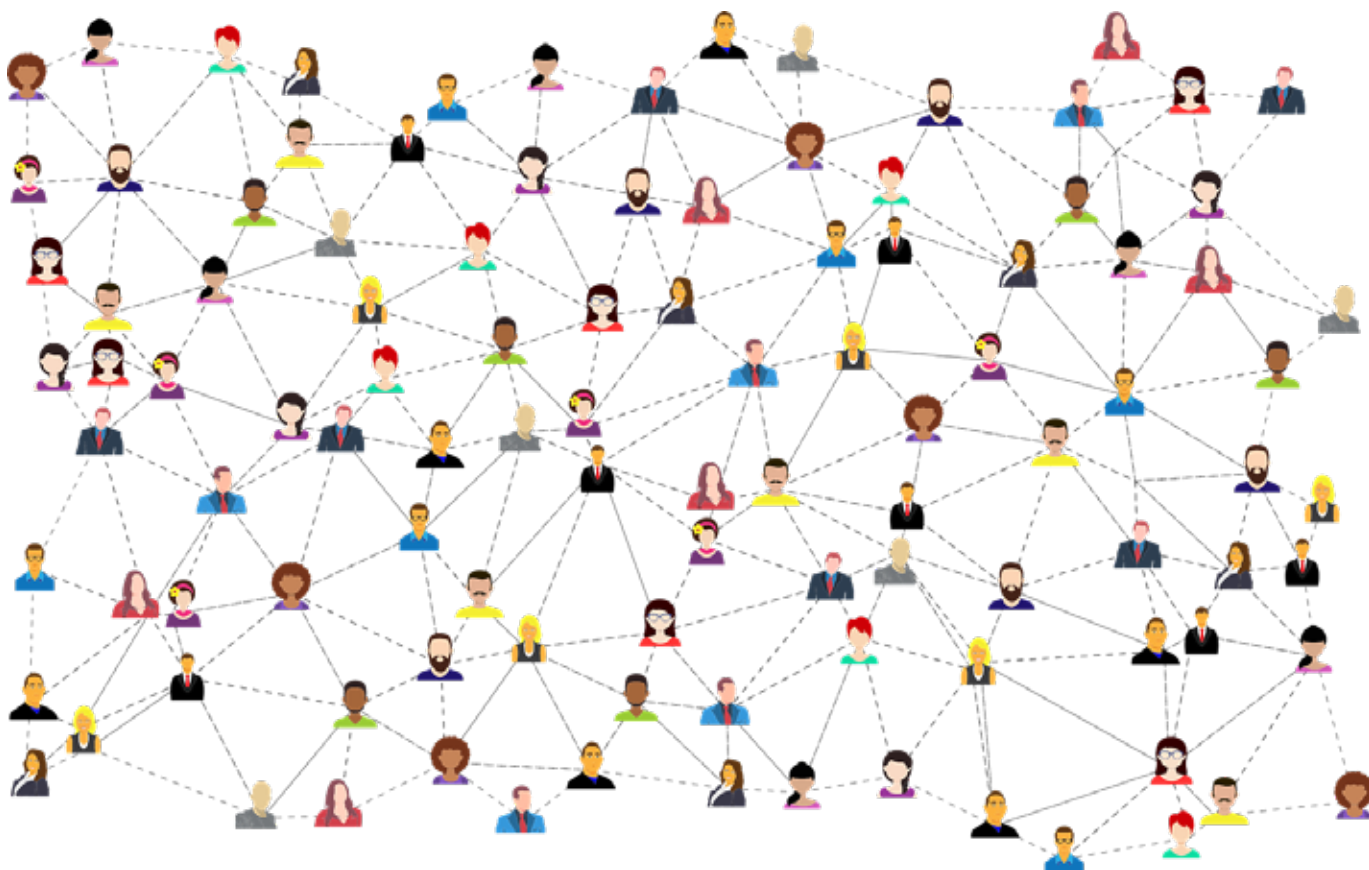
Perspectivas para el futuro.

Llorente admite claramente que “la *maquina* nunca va a escribir mejor que el humano” y que “no ha nacido para sustituir a los humanos, sino para acompañarlos”, es decir los periodistas harán lo que las máquinas no pueden hacer (opinión, reportajes, etc.). Además, coincide con Xavier Ortuño, director adjunto del diario Sport, en que el futuro será una fórmula “*híbrida*” donde “humanos y máquinas convivirán”.

Mejoras

El CEO de *Narrativa* coincide con el director adjunto del Sport ya que admite que las noticias generadas automáticamente tienden a parecerse mucho y destaca que cuantos más datos se inserten, más “la máquina aprenderá a escribir de forma diferente”. Respecto a la falta de “emoción”, también destacada por el periodista, Llorente confirma que la máquina, debidamente entrenada puede aprender. Sin embargo, tratándose de un producto *taylor-made* no compensaría a nivel económico.





Conclusiones

Una vez realizado el estudio, podemos concluir que la generación automática de textos periodísticos es una realidad en aquellos medios que han apostado por incluir la inteligencia artificial para mejorar su servicio de información y a su vez obtener un mayor rendimiento. La diferencia será marcada por aquellos medios que sean capaces de elaborar informaciones únicas para las personas usuarias destinatarias, y este reto se puede conformar con la colaboración entre la plantilla de periodistas y la ayuda de sistemas inteligentes, que ya forman parte del equipo tal y como se ha recogido en los estudios de caso realizados.

Una vez analizado todos los aspectos se puede confirmar que la escritura automática es una suerte de futuro inevitable para el periodismo ya que permite:

- Ahorrar tiempo
- Ahorrar recursos

- Generar más contenido serializable y personalizable
- “Liberar” el periodista que puede dedicarse a tareas más cualitativas

El periodismo tiene a su alcance una cantidad enorme de datos estructurados que estos softwares pueden aprovechar y transformar en noticias, personalizándola y adaptándola a cualquier target.

Sin embargo, tal y como afirma David Llorente, CEO de *Narrativa*, “la *maquina* nunca va a escribir mejor que el humano”.

En otras palabras, la IA será una aliada fundamental para el periodismo, pero solo puede trabajar autónomamente, es decir substituirse al periodista, en la creación de contenidos basados en datos estructurados (finanza, resultados electorales, meteorología, deportes numéricos, etc.). En el mismo deporte, por ejemplo, si la máquina puede generar crónicas técnicamente perfectas de un partido de fútbol, no puede funcionar para deportes como el boxeo donde el factor humano es fundamental.

Por esta razón el futuro parece estar representado por una formula híbrida en la cual la IA proporciona al periodista el “esqueleto” de las noticias (datos básicos, infografía, etc.) para que el periodista lo complete de forma que se pueda adaptar según las diferentes necesidades. Si, tal y como subrayan muchas investigaciones, el futuro parece ser representado por el “*Augmented Human*”, el ser humano aumentado, en este caso, se trataría del “periodista aumentado” capaz, gracias a la ayuda de la IA, de analizar muchos datos y cubrir muchos territorios a la vez, manteniendo el control sobre las tareas de análisis y relacionales (con fuentes, entrevistados, etc.).

Por esta razón es necesario que las empresas cuenten con periodistas especializados en datos, o, en su falta, a transformar sus periodistas en periodistas aumentados, es decir en profesionales acostumbrados a cooperar con las máquinas, ya que, como subrayan los entrevistados, las máquinas “no ha nacido para sustituir a los humanos, sino para acompañarlos”.

...las máquinas
“no ha nacido
para sustituir a
los humanos,
sino para
acompañarlos”...

Referencias

Marco teórico

- Carlson, M. (2015). The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, 3(3), 416–431. <https://doi.org/10.1080/21670811.2014.976412>
- Caswell, D. & Dörr, K. (2018) Automated Journalism 2.0: Event driven narratives, *Journalism Practice*, 12:4, 477-496, DOI:10.1080/17512786.2017.1320773
- Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital journalism*, 3(3), 398-415.
- Dörr, K. N. (2016) Mapping the field of Algorithmic Journalism, *Digital Journalism*, 4:6, 700-722, DOI: 10.1080/21670811.2015.1096748
- Gillespie, Tarleton. (2014). "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot, 167–193. Cambridge, MA: MIT Press.
- Graefe A (2016) *Guide to Automated Journalism*. Tow Center for Digital Journalism Report: 1–48. DOI: 10.1002/ejoc.201200111.
- Haim, M., & Graefe, A. (2017). Automated News. *Digital Journalism*, 5(8), 1044–1059. <https://doi.org/10.1080/21670811.2017.1345643>
- Lewis, S. C. , Guzman, A. L. & Schmidt, T. R. (2019) *Automation, Journalism, and Human-Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News*, *Digital Journalism*, 7:4, 409-427, DOI: 10.1080/21670811.2019.1577147
- Liu, X., Nourbakhsh, A., Li, Q., Shah, S., Martin, R., & Duprey, J. (2017, December). *Reuters tracer: Toward automated news production using large scale social media data*. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 1483-1493). IEEE.
- Lokot, T. & Diakopoulos, N. (2016) News Bots, *Digital Journalism*, 4:6, 682-699, DOI: 10.1080/21670811.2015.1081822

MacCartney, B (2014). Understanding Natural Language Understanding. ACM SIGAI Bay Area Chapter Inaugural Meeting. <https://nlp.stanford.edu/~wcmac/papers/20140716-UNLU.pdf>

Monti, M. (2019). Automated journalism and freedom of information: Ethical and juridical problems related to AI in the press field. *Opinio Juris in Comparatione*, 1.

Nichols, N. (2017) "Natural Language Processing vs. Natural Language Generation". <https://medium.com/@narrativesci/natural-language-processing-vs-natural-language-generation-1b2d18dd0b67>

Shangyuan Wu, Edson C. Tandoc Jr. & Charles T. Salmon (2019). Journalism Reconfigured, *Journalism Studies*, 20:10, 1440-1457, DOI: 10.1080/1461670X.2018.1521299

van Dalen, A. (2012). The algorithms behind the headlines, *Journalism Practice*, 6:5-6, 648-658, DOI: 10.1080/17512786.2012.667268

Veel, K. (2018). Make data sing: The automation of storytelling. *Big Data & Society*. 5(1), 1-8. <https://doi.org/10.1177/2053951718756686>

Yao, L., Peng, N., Weischedel, R., Knight, K., Zhao, D., & Yan, R. (2019, July). Plan-and-write: Towards better automatic storytelling. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 7378-7385).



Webgrafía

The Washington Post (2017) *The Washington Post leverages automated storytelling to cover high school football*. Disponible en: <https://www.washingtonpost.com/pr/wp/2017/09/01/the-washington-post-leverages-heliograf-to-cover-high-school-football/>

Sitio web oficial de Automated insights (2019) Disponible en: <https://automatedinsights.com/customer-stories/associated-press/>

Quill (2020) Disponible en: <https://narrativescience.com/products/quill/>

Editor (2015) Disponible en: <http://nytlabs.com/projects/editor.html>

Vizual.ai (2020) Disponible en: <https://vizual.ai/>

