



---

**Universidad de Valladolid**

**CURSO 2017-2018**

**Facultad de Filosofía y Letras**

**Grado en Periodismo**

**La inteligencia artificial aplicada  
al periodismo: infografías  
automatizadas y narrativa robot**

**Alumna: Paloma Aguado Carro**

**Tutora: Eva Campos Domínguez**

## *Agradecimientos*

*A Eva, por descubrirme. Por su apoyo incondicional y su profesionalidad impecable.*

*A mi familia y amigos, por darme la mano en este precioso camino.*

# **La inteligencia artificial aplicada al periodismo: infografías automatizadas y narrativa robot**

## **Resumen:**

Este Trabajo de Fin de Grado (TFG) tiene como objetivo explorar las aplicaciones de la inteligencia artificial en el periodismo a través del trabajo de generación automática de narrativas de la *startup* Narrativa y del proyecto de generación automática de infografías materializado por la empresa Prodigioso Volcán en colaboración con Software Intelygenz. La elaboración de la investigación parte de la observación participante, con el objetivo de explorar el funcionamiento de ambos proyectos, y entrevistas en profundidad a profesionales especializados en métodos de inteligencia artificial, con la finalidad de conocer las posibilidades de futuro de este periodismo en el contexto español. Las conclusiones del trabajo revelan el uso de estas técnicas por parte de medios digitales españoles, al mismo tiempo que vislumbra el desencanto de las empresas periodísticas con la generación automática de contenidos.

## **Palabras Clave:**

Periodismo robot, noticias e infografías automatizadas, aprendizaje computacional.

**Abstract:**

This Final Degree Project (TFG) aims to explore the applications of artificial intelligence in journalism through the work of automatic narrative generation of the startup Narrativa and the project of automatic generation of infographics materialized by the company Prodigioso Volcán in collaboration with Software Intelygenz. The elaboration of the research starts from the participant observation, with the objective of exploring the functioning of both projects, and in-depth interviews with professionals specialized in artificial intelligence methods, with the purpose of knowing the future possibilities of this journalism in the context Spanish. The conclusions of the research reveal the use of these techniques by Spanish digital media, at the same time that it glimpses the disenchantment of the journalistic companies with the automatic generation of contents.

**Keywords:**

Robot journalism, news and computerized infographics, machine learning

# Índice

|  |           |
|--|-----------|
| <b>1. Introducción .....</b>   | <b>5</b>  |
| <b>2. Estado de la cuestión.....</b>   | <b>6</b>  |
| <b>3. Metodología .....</b>  | <b>10</b> |
| 3.1. Objetivos .....   | 10        |
| 3.2. Hipótesis .....   | 11        |
| 3.3. Tipo de investigación, técnica y enfoque.....   | 13        |
| 3.4. Instrumentos de análisis y recogida de información.....   | 15        |
| 3.5. Selección de la muestra.....  | 16        |
| <b>4. Resultados.....</b>  | <b>18</b> |
| 4.1. Evolución de las técnicas de inteligencia artificial en España: el periodismo como un nuevo campo de introspección.....     | 18        |
| 4.2. El aprendizaje automático en la creación de nuevas narrativas .....   | 21        |
| 4.2.1. El <i>big data</i> como base para la construcción de formatos automatizados: la técnica de <i>matching learning</i> ..... | 22        |
| 4.2.2. Creación de noticias automatizadas a través de la <i>startup</i> Narrativa.....   | 23        |
| 4.2.3. Creación de infografías automatizadas a través de Prodigioso Volcán. ....   | 28        |
| 4.3. La relación del periodista tradicional con las técnicas de inteligencia artificial.....                                     | 32        |
| 4.4. Los límites de la inteligencia artificial: dónde el periodista puede llegar.....  | 34        |
| <b>5. Conclusiones .....</b>   | <b>35</b> |
| <b>6. Bibliografía .....</b>   | <b>40</b> |

|   |           |
|---|-----------|
| <b>7. Anexos .....</b>                              | <b>45</b> |
| 7.1. Ficha a rellenar durante la entrevista .....   | 45        |
| 7.2. Bloques de contenidos de las entrevistas ..... | 46        |
| 7.3. Preguntas de las entrevistas .....             | 47        |

## 1. Introducción

Esta investigación tiene como objeto de estudio analizar la automatización de mensajes periodísticos en España. En concreto, describe cómo trabajan dos empresas españolas en diferentes proyectos para crear contenidos periodísticos a través de inteligencia artificial e identificar el desarrollo de estas dinámicas en el sector mediático en nuestro país. En este sentido, intenta contribuir al estudio de este ámbito del periodismo, novedoso e incipiente en el contexto español, para vislumbrar cuáles son los principales retos a los que se enfrenta la profesión en el presente y futuro. Ramón Salaverría (2014) ya apuntaba hace cuatro años la incidencia del *robot journalism* en las dinámicas periodísticas. En España aún no adquiriría una gran relevancia, pero sí en los países anglosajones, donde las investigaciones acerca de esta nueva forma de hacer un periodismo más rápido y económico a través de un contenido generado automáticamente en grandes cantidades, están en continua evolución. Esta nueva dinámica de crear contenido periodístico se conoce como *robot journalism* (Van Dalen, 2012) y es el objeto de estudio principal del presente Trabajo de Fin de Grado.

Los ejemplos de periodismo computacional o *robot journalism* desarrollados en otros países, como Gran Bretaña, Francia o Estados Unidos, son el principal incentivo para revelar el uso de estas técnicas en España y tratar de esclarecer su incidencia en el sector periodístico. El caso pionero de la agencia de noticias estadounidense *Associated Press* en la generación de noticias deportivas y económicas de forma automatizada a través de una herramienta de software desarrollada por *Automated Insights* (Dörr y Hollnbuchner, 2016), se establece como referente en otros países en el desarrollo de inteligencia artificial con aplicaciones para el periodismo. En el 2014, la agencia de noticias publicaba 4.000 historias de forma automática frente a las 300 que realizaban los periodistas (Sánchez y Sánchez, 2017, p.4). En Europa, *The Guardian* también se ha configurado como uno de los medios de comunicación pioneros en

incorporar estas técnicas a su redacción en el año 2010, a través de *bots* capaces de construir historias financieras, deportivas e incluso; proponer temas a través de métodos de inteligencia artificial (Sánchez y Sánchez, 2017). La existencia y la aplicación de las técnicas de Inteligencia Artificial en otros países, al igual que las investigaciones realizadas en esta área, son el principal motivo para desarrollar el presente Trabajo de Fin de Grado y así, poder esclarecer su uso en España.

## **2. Estado de la cuestión**

La labor periodística desarrollada en los medios de comunicación experimenta, desde el pasado siglo, una continua introspección hacia nuevos modelos y formatos. La revolución tecnológica es el pilar fundamental de esta búsqueda incesante por definir qué significa ‘hacer periodismo’ en nuestros días (Salaverría y García, 2008). Una de las transformaciones más imponentes para el sector se ha materializado a través de la aplicación de métodos de inteligencia artificial a las dinámicas tradicionales del periodista. Es decir, la capacidad de replicar la estructura de los textos periodísticos de forma automática a través de una serie de algoritmos programados (Bunz, 2010).

Esta reciente incorporación de la tecnología robótica al periodismo provoca una confusión terminológica de los diferentes autores que han comenzado a investigar este campo. Por esta razón, es conveniente señalar que existen varias formas de definir el término periodismo robot o *robot journalists* (Van Dalen, 2012 y Latar 2015), asociado a la creación de contenido automatizado a través del uso de nuevas tecnologías (Clerwall, 2014). Así, encontramos otros términos semejantes como periodismo computacional (Clerwall, 2014) o periodismo algorítmico (Berobici, 2010). En definitiva, todos ellos hacen referencia a una nueva narrativa, que surge a raíz de una constante evolución tecnológica en el campo periodístico. Por lo tanto,



las tareas que antes solo podían realizar los periodistas tradicionales, ahora cuentan con una tecnología algorítmica que puede organizar la información y convertirla en relatos o contenidos periodísticos en un periodo de tiempo muy escaso (Latar, 2015).

El presente trabajo, a pesar de ser una investigación centrada en el área de las ciencias sociales, también necesita de la explicación de términos relacionados con la informática y la ciencia, que afectan al entendimiento del concepto *robot journalism*. De esta forma, los conceptos explicados en el presente trabajo se engloban en un reciente área de conocimiento denominado ciencia social computacional (Lazer, 2009). Uno de los conceptos más importantes a la hora de alcanzar esta comprensión son los *bots o robots* aplicados en medios de comunicación, que se definen como herramientas que sirven para realizar una tarea periodística de forma automática, sin la colaboración de un ser humano. La primera vez que se utilizó el término ‘robot’ en los medios de comunicación fue en 1998, cuando Lee y Kim propusieron un sistema de noticias personalizado para el usuario a través de un algoritmo (Dörr y Hollnbuchner, 2016). Así, una de las primeras aplicaciones de los *bots* de noticias está relacionada con la distribución de las mismas a través de diferentes plataformas digitales (Cervera, 2017). Hablamos, por lo tanto, de aplicaciones conversacionales que permiten al usuario consumir una información filtrada y personalizada (Sánchez y Sánchez, 2017). De esta forma, no solo las herramientas surgidas de la evolución tecnológica sirven al periodista para crear contenido, también para su distribución a la audiencia objetivo. El concepto *robot journalism*, puede, como se ha explicado con anterioridad, definirse con otros términos como periodismo algorítmico (Berobici, 2010). En este sentido, es necesario explicar el significado de un algoritmo, como un método con diferentes aplicaciones que permite solucionar un problema paso a paso (Terminus7 y Intelygenz Academy Teams, 2018, pp7).

Del mismo modo, también es importante señalar las técnicas de *big data* como un pilar fundamental para comprender el concepto de periodismo robot. El *big data* constituye una técnica de recopilación y estructuración de grandes cantidades de datos. Es, por tanto, la base principal para, después, aprovecharlos y organizarlos a través de diferentes técnicas de automatización (Lewis y Westlund, 2014). En este sentido, la recopilación de grandes cantidades de datos adquiere relevancia cuando se aplican determinadas técnicas para extraer información y resultados computables. Así, hablamos de *data mining* o minería de datos, término referido a la capacidad de extraer información, que derive en un posterior conocimiento, a través de las relaciones entre grandes cantidades de datos (Arcila-Calderón, Barbosa-Caro y Cabezuelo-Lorenzo, 2016). Este concepto comenzó a tener relevancia en la década de los noventa, donde se posicionaba como una nueva herramienta aplicable a diferentes sectores profesionales y dirigida al aprovechamiento de estos datos para explicar determinados comportamientos y realizar predicciones de futuro (Dhar, 2012). Además, cabe destacar el término anglosajón *machine learning comprehension* –traducido como algoritmos de aprendizaje automático– como base principal para comprender en qué se basa una de las corrientes más actuales del *robot journalism* (Burrell, 2016). Esta técnica, descrita como un subconjunto de la inteligencia artificial, se basa en proporcionar datos a un ordenador para que este pueda aprender por sí mismo, y de esta forma, realizar el trabajo sin la colaboración humana a partir de experiencias (Terminus7 y Intelygenz Academy Teams, 2018, pp7).

El fundador de Narrative Science –una de las empresas pioneras en el desarrollo de contenido automatizado– asegura que esta información creada a través de algoritmos es solo el comienzo de un futuro universo de noticias generadas automáticamente. Es decir, el periodismo robot presenta un futuro, todavía, repleto de posibilidades. Algunos investigadores señalan al periodismo robot como una nueva forma de liberación de tareas superfluas o menos relevantes

para el periodista tradicional (Bunz 2010). Así, las tareas más importantes en un trabajo periodístico de alta calidad, como la creatividad o la sensibilidad, recaerán sobre la figura del periodista, mientras los trabajos más rutinarios relacionados con la reunificación y el ordenamiento de datos, serán realizados por las máquinas (Van Dalen, 2012). Por lo tanto, existirá una clara diferencia entre los contenidos periodísticos creados de forma artificial y aquellos elaborados por el ser humano. Sin embargo, estudios como el realizado por Christer Clerwall (2014), aseguran que la audiencia no identifica amplias diferencias entre los textos realizados por periodistas y los creados a través de inteligencia artificial. Por esta razón, sí que podría suponer una amenaza para el trabajo del periodista tradicional, que podría verse sustituido en sus funciones más habituales (Van Dalen, 2012).

De esta forma, existirían dos corrientes diferenciadas a cerca de los cambios que producirá la incorporación de las técnicas de inteligencia artificial en el ámbito periodístico. Dos caminos que confluyen en considerar la llegada de la inteligencia artificial a las redacciones de los medios un cambio sin precedentes, que obligará a los periodistas a centrarse en trabajos poco rutinarios y desarrollar otro tipo de habilidades más creativas que aporten valor al trabajo periodístico (Ford, 2013). En cualquier caso, las aplicaciones de la inteligencia artificial en el área de las ciencias comunicativas trascienden el cómputo de datos y su transformación a una narrativa legible. Pueden aplicarse, incluso, para generar nuevas ideas mecedoras de la investigación del periodista a través del análisis de datos (Broussard, 2014).

La introspección hacia esta nueva forma de ‘hacer periodismo’ se está realizando de forma paralela a los medios de comunicación, en muchas ocasiones a través de *startup* periodísticas que apuestan por descubrir los nuevos formatos (Valero-Pastor, González Alba, 2018). Las empresas periodísticas, por tanto, conviven en el mismo entorno que estas entidades y según autores como Gynnild (2014), existe confusión en los medios de comunicación a la hora de

determinar las nuevas posibilidades que ofrece la tecnología más avanzada. Por esta razón, el presente Trabajo de Fin de Grado pretende analizar las relaciones existentes entre las nuevas técnicas más avanzadas de inteligencia artificial en España, los medios de comunicación y las empresas o *startups* encargadas de desarrollar dichas técnicas.

### 3. Metodología

En el presente apartado del trabajo se explica la metodología de la investigación. En primer lugar, se presentan los objetivos, hipótesis y preguntas de investigación. En segundo lugar, se describe el tipo de investigación y la técnica utilizada. Finalmente, se muestran los instrumentos y técnicas utilizados en la investigación y la muestra seleccionada.

#### 3.1. Objetivos

Con el objetivo de indagar en las aplicaciones del periodismo robot, se analizan dos proyectos concretos de *robot journalists* en España, de diferente tipología, con la finalidad de explorar sus características e implantación. Por un lado, la creación de infografías a través de un proyecto dirigido por la empresa Prodigioso Volcán; y por otro, la actividad que lleva a cabo la *start-up* Narrativa en relación a la creación de contenido automatizado. Como objetivos secundarios de la investigación se establecen los siguientes:

- **Objetivo 1:** Explorar el comportamiento de los medios de comunicación ante las propuestas estudiadas, con el objetivo de determinar su adaptación a las nuevas corrientes tecnológicas.
- **Objetivo 2:** Conocer las limitaciones de las técnicas de Inteligencia Artificial aplicadas al periodismo y a medios de comunicación, para así determinar las funciones que puede desempeñar únicamente un periodista.

- **Objetivo 3:** Descubrir las principales relaciones laborales entre las empresas especializadas en técnicas de inteligencia artificial con las empresas periodísticas.
- **Objetivo 4:** Indagar en los cambios principales en las redacciones que derivan de la automatización de determinadas funciones periodísticas, a través de las declaraciones de expertos y de las experiencias de las empresas que desarrollan las técnicas de automatización.

Se establecen, así, una serie de preguntas de investigación: ¿los proyectos relacionados con la inteligencia artificial son relevantes en el sector periodístico?, ¿Los medios de comunicación tradicionales muestran interés por conocer estas dinámicas para aplicarlas en sus redacciones o por el contrario, se muestran reticentes?, ¿debe existir una colaboración necesaria entre medios de comunicación y empresas o *startups* dedicadas a la innovación tecnológica en materia de inteligencia artificial?, ¿está teniendo consecuencias directas en el presente de la profesión periodística?, ¿hasta qué punto pueden afectar las nuevas dinámicas al futuro del periodismo?

### 3.2. Hipótesis

Teniendo en cuenta los objetivos expuestos y las preguntas de investigación planteadas, se pretenden confirmar o refutar las siguientes hipótesis:

- **Hipótesis 1:** El periodismo automatizado está adquiriendo una mayor relevancia en el panorama nacional a través de *startups* que apuestan por ofrecer noticias u otro tipo de contenidos generados automáticamente a los medios de comunicación tradicionales.

Las *startups*, por su estructura flexible, pueden dedicar mayor tiempo a la investigación e innovación de nuevas técnicas en diferentes áreas (Valero-Pastor y González-Alba, 2014). Sin

embargo, el papel de los medios de comunicación es fundamental, puesto que tienen que hacer un esfuerzo por comprender las nuevas técnicas aplicables a los métodos tradicionales de generación de noticias (Gynnild, 2013).

- **Hipótesis 2:** El periodismo automatizado tendrá mayor incidencia en las informaciones basadas en datos o estadísticas, sin embargo, no modificará otras informaciones generadas a través de otras técnicas no relacionadas con el *big data*.

La generación de contenido automatizado necesita disponer de grandes cantidades de datos para establecer relaciones y estructurar el contenido (Clerwall, 2014). Por esta razón, las primeras noticias e informaciones generadas de forma automática versan sobre temática deportiva o financiera –existen una gran cantidad de datos disponibles en abierto– como las publicadas por la agencia de noticias *Associated Press* (Dörr y Hollnbuchner, 2016).

- **Hipótesis 3:** En la actualidad es imprescindible, todavía, la intervención del periodista en la generación de informaciones, a pesar de utilizar técnicas de inteligencia artificial.

El uso de técnicas de inteligencia artificial libera al periodista de las tareas más rutinarias y brinda al profesional tiempo para dedicarse a un trabajo periodístico de alta calidad. Por lo tanto, solo una parte del proceso de generación de información –el trabajos más rutinarios relacionados con la reunificación y el ordenamiento de datos– serán realizados a través de inteligencia artificial (Van Dalen, 2012).

- **Hipótesis 4:** Los medios de comunicación se muestran reticentes a incorporar nuevos modelos de generación automática de contenido –a través de textos, imágenes o infografías– por temor a una desestructuración de las dinámicas tradicionales y pérdida de puestos de trabajo.

José Cervera (2017) asegura que “las técnicas del *big data*, los sistemas de inteligencia artificial basados en *deep learning* y la existencia de internet, la mayor biblioteca accesible del planeta, están permitiendo la creación de sistemas capaces de complementar, ayudar e incluso (en ciertas limitadas circunstancias) reemplazar a los periodistas en sus tareas cotidianas” (p.2). Por esta razón, los medios de comunicación pueden apostar por una postura de distanciamiento ante los cambios que pueda generar en sus dinámicas tradicionales.

### **3.3. Tipo de investigación, técnica y enfoque**

Se plantea un estudio exploratorio basándose en entrevistas en profundidad y observación participante. El enfoque cualitativo permitirá conocer de primera mano una realidad cambiante y novedosa con una mayor flexibilidad, necesaria para afrontar los imprevistos que puedan surgir a raíz de su aplicación al campo de investigación descrito anteriormente (Bryman, 1984). De esta forma, y con el propósito de cumplir los objetivos propuestos, se han seleccionado dos métodos de enfoque cualitativo, que responderán a la naturaleza de esta investigación.

En primer lugar, la observación participante se establece por varios autores como un pilar fundamental de la metodología cualitativa, pero sin embargo, se muestra aún en continuo desarrollo y provoca el enfrentamiento de los teóricos a la hora de definir sus principales puntos fuertes (Callejo, 2002). Por otra parte, la observación participante también se configura como uno de los principales métodos utilizados en las ciencias sociales en continua evolución, que permite al investigador adentrarse en dinámicas desconocidas e informar sobre las mismas de forma exhaustiva (Kawulich, 2005). En el presente Trabajo de Fin de Grado se realizan dos observaciones participantes, encaminadas a responder a los objetivos propuestos. En primer lugar, se asistirá a una jornada de trabajo de la empresa Narrativa, concretamente a un encuentro programado para establecer las pautas de acción de un nuevo proyecto de generación

automática de noticias en las próximas elecciones municipales y generales. En segundo lugar, la empresa Prodigiso Volcán mostrará el *modus operandi* del primer prototipo de generación automática de infografías.

Del mismo modo, se realizarán dos entrevistas en profundidad sistematizadas, con preguntas abiertas y estructuradas, con el objetivo de conocer la realidad del objeto de la investigación de la forma más exhaustiva. Con el propósito de no sesgar la investigación por los intereses personales de la investigadora, la observación participante será posterior a la entrevista en profundidad realizada a Ramón López de Mántaras, el director del Instituto de Investigación de Inteligencia Artificial, y a Pablo Gervás Gómez-Navarro, director del grupo de investigación de la Universidad Complutense de Madrid Procesamiento de Lenguaje Natural. Así, se pretende establecer un primer análisis descriptivo de la situación actual de los métodos de inteligencia artificial aplicados a los medios de comunicación, y de esta forma, poder desarrollar una posterior observación activa encaminada a comprobar los datos extraídos en este primer contacto (Bernard y Gravlee, 2014).

En primera instancia se elaborarán las estrategias metodológicas a seguir en la entrevista en profundidad a través de unas preguntas programadas que respondan a los objetivos de la investigación y que además, sirvan de base para realizar las posteriores observaciones participantes. De esta forma, podrán modificarse posteriormente determinados aspectos relacionados con el programa de la observación participante con el objetivo de comprobar las hipótesis planteadas después de conocer, más a fondo, el estado actual de los métodos de inteligencia artificial en España y su repercusión en el periodismo. Por último, tendrá lugar el contraste de las informaciones extraídas tras la observación participante y las entrevistas en profundidad, que se redactarán después de finalizar el estudio. Este proceso sigue los pasos del



planteado por Janesick (1994), quien dividen en tres fases el diseño de la investigación cualitativa: el principio del estudio, su desarrollo y la redacción final del mismo.

Los resultados, por lo tanto, responderán a los objetivos planteados y se estructurarán, de esta forma, en tres apartados. En primer lugar, la información relativa al desarrollo de las técnicas de inteligencia artificial en España, a través de los datos obtenidos de las entrevistas en profundidad anteriormente mencionadas. En segundo lugar, se profundizará en el proceso de creación de las narrativas automatizadas y formatos generados a través de inteligencia artificial, con el objetivo de vislumbrar la existencia de una repercusión en el ámbito mediático y periodístico. Por último, se establecerán las principales limitaciones de los métodos de automatización de contenidos, de forma que se vislumbre la relación del periodista con las técnicas de inteligencia artificial.

### **3.4. Instrumentos de análisis y recogida de información**

La presente investigación necesitará de una serie de instrumentos de análisis para la recogida de la información necesaria para cumplir con los objetivos del presente Trabajo Fin de Grado. Así, la entrevista sistematizada y el cuaderno de campo serán las dos herramientas indispensables para el desarrollo de esta investigación. El cuaderno de campo estará presente durante todo el proceso de interacción con la empresa Prodigioso Volcán y la *startup* Narrativa, con el objetivo de recoger todos los datos permitentes (Kawulich, 2005). Las entrevistas en profundidad, previas a las observaciones participantes, estarán estructuradas en bloques temáticos con preguntas sistematizadas.<sup>1</sup>

---

<sup>1</sup> Ver anexo 7.2. y 7.3. “Bloques de contenidos de las entrevistas y preguntas de las entrevistas”

### 3.5. Selección de la muestra

La muestra seleccionada para este trabajo pretende vislumbrar nuevas formas de creación de contenido periodístico a través de la automatización. Por un lado, se estudiará un proyecto concreto de la empresa Prodigioso Volcán, en colaboración con la empresa Software Intelygenz; y por otro, se analizará el proceso de creación de narrativas automatizadas de la *startup* Narrativa, a través del software ‘Gabriele’ que han realizado en colaboración con la Universidad de Alcalá de Henares. Ambos proyectos se han seleccionado con el fin de representar dos realidades diferentes de información periodística. En primer lugar, la creación de infografías o imágenes visuales a través de procesos de *machine learning comprehension* y en segundo lugar, la creación de textos generados a través de esta misma técnica. Así, distinguimos dos realidades diferentes que responden a necesidades actuales en los medios de comunicación.

La *startup* Narrativa y la empresa Prodigioso Volcán se han seleccionado como la muestra principal, debido a su relevancia en el ámbito nacional y la vinculación directa de sus proyectos con el desarrollo de técnicas de inteligencia artificial aplicadas al periodismo.

Narrativa es la única empresa en la actualidad que ofrece un servicio de generación automática de noticias en España a medios de comunicación. Por lo tanto, es indispensable su labor para esclarecer las implicaciones de la inteligencia artificial en el periodismo. Además, no solo trabaja con medios españoles, por lo que puede aportar una visión más global y una comparativa entre los servicios que ofrecen en los tres países en los que opera.

Por otro lado, Prodigioso Volcán es una empresa de consultoría estratégica, diseño y generación de contenidos. Es, en la actualidad, una de las mil empresas que más crece en

Europa según *Financial Time*.<sup>2</sup> Para la realización de esta investigación, el interés principal reside en el proyecto de inteligencia artificial de generación de infografías automatizadas financiado por Google DNI Fund.<sup>3</sup> Prodigioso Volcán es, así, una de las tres empresas en España que han obtenido un proyecto prototipo de innovación de esta iniciativa europea. Además, en ambos casos, se llevarán a cabo entrevistas a los dirigentes de los proyectos. En el caso de Narrativa, a David Llorente, CEO y fundador de la empresa; y en Prodigioso Volcán, a Ana Ormaechea.

Por último, un factor de interés para el presente Trabajo Fin de Grado reside en la comparación de las dinámicas de ambas entidades, puesto que Narrativa se define como una *startup*<sup>4</sup> y Prodigioso Volcán suma ocho años de experiencia.

Las entrevistas en profundidad realizadas con anterioridad a las observaciones participantes se han realizado teniendo en cuenta las necesidades de la investigación y la trayectoria profesional y académica de los entrevistados (López y Deslauriers, 2011). Las entrevistas, estructuradas en bloques temáticos, se desarrollaron durante una hora en los respectivos centros de investigación. En el caso de realizada a Ramón López de Mántaras, en el Instituto de Investigación de Inteligencia Artificial de Barcelona, y a Pablo Gervás Gómez-Navarro en la Universidad Complutense de Madrid Procesamiento de Lenguaje Natural.

---

<sup>2</sup> Información extraída del informe de *Financial Times*: <https://ig.ft.com/ft-1000/2018/> [Consultado el 27/06/18]

<sup>3</sup> Google DNI Fund es un programa europeo que forma parte de Google News Initiative, un proyecto que trata de incentivar la innovación en el ámbito periodístico. En este caso, Prodigioso Volcán decidió presentar un proyecto de generación automática de infografías a través de métodos de inteligencia artificial y *machine learning*, que, posteriormente, aprobó Google DNI Fund. Información extraída de: <https://newsinitiative.withgoogle.com/dnifund/> [Consultado el 27/06/18]

<sup>4</sup> Una *startup*, según BBVA es una “empresa emergente, normalmente con un alto componente tecnológico, con grandes posibilidades de crecimiento y que, por lo general, respalda una idea innovadora que sobresale de la línea general del mercado”. Información extraída de <https://www.bbva.com/es/que-es-una-startup/> [Consultado el 27/06/18]

Las conclusiones extraídas de estas dos experiencias serán extrapolables al universo investigado, estableciendo una serie de pautas para comprender cómo se desarrollan iniciativas de inteligencia artificial en España (Mejía, 2000). Así, para comprender el desarrollo del trabajo y mostrar de forma concisa los resultados obtenidos, se plantea una tabla con los principales resultados obtenidos de la observación participante y la estructuración de los mismos en relación a las hipótesis planteadas.

#### **4. Resultados**

En este apartado se exponen los resultados obtenidos a raíz de la aplicación de la metodología descrita. En el primer apartado se analiza el estado actual de las técnicas de inteligencia artificial en España y su repercusión en el sector periodístico; en segundo lugar, se profundiza en el funcionamiento y las dinámicas de los dos proyectos seleccionados, además de incidir en las técnicas de *big data* en los medios. Finalmente, se tratará la relación del periodista tradicional con las técnicas de inteligencia artificial y sus límites de aplicación.

##### **4.1. Evolución de las técnicas de inteligencia artificial en España: el periodismo como un nuevo campo de introspección**

Los entrevistados coinciden en señalar que las técnicas de inteligencia artificial son una realidad presente en los medios de comunicación en España. Sin embargo, no es el principal campo de investigación y aplicación, debido a una falta de demanda latente en los medios tradicionales (Mántaras, 2017). Del mismo modo, señalan la falta de inversión como el principal problema al que se enfrentan los medios de comunicación para incorporar en sus redacciones nuevas técnicas de generación automática de contenidos. En este sentido, otros campos como el médico, financiero o el del entretenimiento, acaparan un mayor número de investigaciones por la creciente demanda en estos sectores.

Este hecho se materializa en el escaso número de investigaciones realizadas por parte de entidades públicas de investigación, como la Universidad Complutense de Madrid, o centros de investigación privados, como el Instituto de Investigación de Inteligencia Artificial (Barcelona). En el primer caso, el grupo de investigación dirigido por Pablo Gervás Gómez-Navarro de Procesamiento de Lenguaje Natural, asegura no haber realizado ningún trabajo en colaboración con medios de comunicación. Sin embargo, sí han mantenido relación con la *Cadena Ser* para llevar a cabo un proyecto de generación automática de preguntas para una entrevista. En palabras del entrevistado, el proyecto trataba de “generar una lista de preguntas a través de la información que previamente recibía el sistema del entrevistado”. Sin embargo, no ha llegado a ponerse en práctica y se ha paralizado. En el caso del Instituto de Investigación en Inteligencia Artificial (IIIA), dirigido por Ramón López de Mántaras, únicamente han llevado a cabo un proyecto de inteligencia artificial para un medio de comunicación. En este caso, se trata de un proyecto en colaboración con Buscamedia para desarrollar una programación automática de narrativas audiovisuales (*Automatic Generation of Audiovisual Narrative*). Este proyecto finalizó en 2011 y se aplicó en TV3 (Televisión de Cataluña). El proyecto funcionaba de la siguiente forma, según explica el entrevistado Ramón López de Mántaras:

En una competición deportiva están los periodistas grabando el acontecimiento con las cámaras de televisión. Pero también, cuando acontece algo importante como un penalti, un gol o una falta (algo que a criterio del periodista es relevante) realizan una anotación manual del minuto en el que ha sucedido. En base a estos datos que genera el periodista *in situ* acompañados de la grabación del partido, el sistema es capaz de hacer un resumen de un minuto en muy poco tiempo para mostrarlo en el telediario. El sistema decide en base a las anotaciones y a las imágenes cuales son las secuencias que

tiene que mostrar durante un minuto (...) Además, tiene de referencia otras carreras y el resumen de las mismas, de forma que le sirve para replicar este procedimiento.

Del mismo modo, ninguno de los entrevistados dirige proyectos relacionados directamente con medios de comunicación. Sin embargo, sí señalan la incidencia de las nuevas técnicas de inteligencia artificial en el periodismo como un factor a tener en cuenta en los próximos años. De hecho, esta repercusión se muestra a través del surgimiento de *startups* en España relacionadas con métodos de inteligencia artificial y software (Mántaras, 2017). De igual forma, sitúan a España como un país con un buen desarrollo de técnicas de innovación aplicadas a inteligencia artificial. Según el entrevistado Ramón López de Mántaras, cerca de 700 personas están trabajando en la actualidad en este campo.

Por otro lado, es conveniente señalar, a raíz de las observaciones realizadas a las empresas Prodigioso Volcán y Narrativa, el desconocimiento y el recelo por parte de las grandes empresas mediáticas como principales razones del distanciamiento de las técnicas de generación automática de contenidos. En el caso de la empresa Prodigioso Volcán, el proyecto a análisis de creación de infografías de forma automatizada no ha sido presentado a los medios de comunicación en la fecha de entrega de este Trabajo Fin de Grado. Sin embargo, la codirectora del mismo, Ana Ormaechea, asegura que los medios muestran reticencias al uso de técnicas automatizadas y es por esta razón, por la que Prodigioso Volcán decidió comenzar con un proyecto de inteligencia artificial con un funcionamiento simple y de una fácil comprensión para los medios de comunicación. En el caso de Narrativa, David Llorente, el director general de la empresa, asegura que tras su surgimiento como *startup* hace tres años, se encontraron con el distanciamiento de los grandes medios de comunicación. Sin embargo, el entrevistado confiesa que la evolución ha sido positiva y la respuesta de los medios de comunicación está cambiando. Prueba de este hecho es el aumento de los medios de comunicación que incorporan

su método de creación automática de informaciones a sus redacciones. Sin embargo, también cabe destacar que España se encuentra en una posición inferior con respecto a otros países europeos como Alemania o Estonia, y a una gran distancia de otros como Estados Unidos o Emiratos Árabes. De hecho, gran parte de los beneficios obtenidos por Narrativa provienen de los países anteriormente mencionados. En España, al no contar con el apoyo suficiente por parte de los grandes conglomerados mediáticos, investigan nuevas técnicas de financiación a través de la generación automática de contenidos para servicios financieros, el sector sanitario o el comercio electrónico.

En el siguiente apartado, se estudiará con mayor precisión la metodología y los procesos relacionados con esta generación automática de contenidos a través de inteligencia artificial, al igual que se mostraran ejemplos de noticias generadas automáticamente publicadas en la actualidad por medios de comunicación.

#### **4.2. El aprendizaje automático en la creación de nuevas narrativas**

En los dos casos seleccionados, la forma de trabajar y llevar a cabo los diferentes proyectos de inteligencia artificial responde a un modelo de colaboración y espíritu emprendedor. En primer lugar, la empresa Prodigioso Volcán realiza el proyecto en colaboración con la entidad Intelygenz Software. De esta forma, a pesar de ser Prodigioso Volcán la encargada de presentar el proyecto a Google DNI Found y de explicar el prototipo de generación automática de contenidos a los medios de comunicación, es Intelygenz Software la encargada de elaborar proceso de automatización a través de las técnicas de *matching learning*. Esta labor necesitan llevarla a cabo de forma conjunta, ya que Prodigioso Volcán no es una empresa especializada en técnica de inteligencia artificial ni tiene las herramientas necesarias para poder desarrollar el proyecto completo. En segundo lugar, la *startup* Narrativa funciona a través de una cátedra

en la Universidad Alcalá de Henares (Madrid) con la colaboración de investigadores en el campo de la inteligencia artificial. El grupo está compuesto de forma íntegra por ingenieros, a pesar de contar con la colaboración de periodistas para determinadas tareas que se analizarán posteriormente. En este caso, la *startup* tiene establecida su sede principal en Estonia y cuenta con otras dos sedes más, en Madrid y en Dubai. Por lo tanto, estamos ante dos modelos empresariales diferentes, aunque con características comunes, ya que ambos trabajan en proyectos de innovación a través de la colaboración con empresas especialistas en el sector o entidades públicas especializadas y tienen un espíritu de emprendimiento, reflejado en la propia evolución de los dos proyectos analizados en el siguiente apartado.

#### **4.2.1. El *big data* como base para la construcción de formatos automatizados: la técnica de *machining learning***

En este punto del trabajo se pretende analizar, a rasgos generales, cómo pueden aplicarse técnicas de inteligencia artificial al trabajo periodístico. Así, después se estudiarán los dos proyectos seleccionados a través de las observaciones participantes.

Los entrevistados señalan las técnicas de *big data* como los cimientos sobre los que asentar el fenómeno de la inteligencia artificial con aplicaciones periodísticas. Ramón López de Mántaras explica como el *big data* “no es que sea una técnica de la Inteligencia Artificial, sino una característica que tiene una base de datos, donde existe gran abundancia de datos que después, son analizados a través del *data mining*”. Por lo tanto, estos datos, después, servirán a los sistemas de inteligencia artificial como base para generar nuevos contenidos. Una de las formas principales para llevar a cabo este proceso es a través del *machining learning*. Las dos experiencias analizadas sustentan su metodología de trabajo en ambos procesos. Los entrevistados aseguran que el *big data* es el comienzo para después realizar *data mining* o



minería de datos, es decir, procesar la información para descubrir patrones repetidos con el objetivo de perseguir un determinado resultado. De esta forma, después, la máquina puede aprender a relacionar estos datos a través de experiencias previas, técnica que recibe el nombre de *machine learning*.

Ramón López de Mántaras así explicaba cómo pueden funcionar estos sistemas:

A veces, en lugar de programar paso a paso todo lo que tiene que hacer un sistema para resolver un problema, se le dan ejemplos de problemas resueltos. Por ejemplo, el sistema recibe ejemplos de partidas de ajedrez, de partidas de *go* o de imágenes médicas. Así, se entrena al sistema para que pueda discriminar si hay un tumor o no (basándose en imágenes médicas) o cuál es la mejor jugada (si lo aplicamos a juegos).

Además, también es de relevancia señalar que los datos con los que ‘se entrena’ a una máquina para que reproduzca determinados comportamientos, deben tener siempre un valor y un sentido real. Es decir, no vale únicamente con tener acceso a grandes cantidades de datos (Gervás, 2018). El objetivo, por lo tanto, reside en conseguir que la máquina sea capaz de detectar patrones repetidos a través de los datos y las experiencias previas, para después, replicarlos y convertirlos en un nuevo material.

#### **4.2.2. Creación de noticias automatizadas a través de la *startup* Narrativa.**

En el presente apartado del trabajo, se explicará cómo se lleva a cabo el proceso de producción de narrativas automatizadas y el papel de la figura del periodista dentro de este mismo proceso. Además, se mostrarán ejemplos de narrativas generadas automáticamente publicadas en la actualidad en medios de comunicación, con el objetivo de analizar la repercusión actual de las técnicas de inteligencia artificial en las empresas periodísticas.

La *startup* Narrativa es la única empresa en el panorama nacional, según su propio CEO, David Llorente, que ofrece en contenidos generados automáticamente en medios de comunicación. En la actualidad, se dedican a la generación de contenidos automáticos a través de un software que han denominado ‘Gabrielle’. El método de trabajo para generar el contenido a través de este software se divide en cuatro fases principales: en primer lugar, existe una parte de ingesta de datos; en segundo lugar, el análisis de los mismos; en tercer lugar, la generación de contenido; y por último, la publicación de la información. Todos los pasos se llevan a cabo de forma automática, sin la necesidad de la intervención de un periodista. Los textos, de la misma forma, no se modifican en ningún aspecto.

En la actualidad, Narrativa genera 12.000 noticias a la semana que se publican en medios de comunicación españoles. La temática de estas noticias siempre es financiera o deportiva. Los clientes actuales son *El Confidencial*, *El Independiente*, *El Español*, el *Grupo Z*, *Vocento* y *MediaPro*. Todos los contenidos generados se publican en el formato web de los diferentes medios. Por lo tanto, medios de comunicación de referencia en España ya cuentan con contenidos generados automáticamente.

## Imagen 1

*Noticia generadas automáticamente por Narrativa y publicadas por El Independiente*



The image is a screenshot of a news article from the website 'EL INDEPENDIENTE'. The article is titled 'Mercados y divisas 15/06/2018' and is categorized under 'ECONOMÍA'. It was published on June 15, 2018, at 10:51. The article text states: 'El metal dorado sufre una leve subida. Si hace dos días se pagaba a 1.294,40 dólares por onza ayer lo hizo a 1.305,30 dólares por onza, lo que supone una leve subida del 0,84%.' There are social media sharing icons for Facebook, Twitter, LinkedIn, and Google+. A 'Cotización Oro 15/06/2018: 1.305,300' is also visible. On the right side, there is a 'TE RECOMENDAMOS' section with a small image of a woman speaking in front of a Spanish and European Union flag.

Fuente: captura de pantalla de <https://www.elindependiente.com/>

## Imagen 2

*Una de las noticias generadas automáticamente por Narrativa y publicadas por El As*



The screenshot shows a news article on the AS.com website. The main headline is "Playoff para el Mestalla y descenso para L'Hospitalet". Below the headline, there is a sub-headline: "El Valencia Mestalla jugará los playoff de ascenso y L'Hospitalet no consigue la salvación y jugará el año que viene en tercera." The article text begins with "Los visitantes llegaban a este encuentro en decimoséptima posición y con necesidad de conseguir la victoria y esperar que el Atlético Levante saliese derrotado de su encuentro contra el Badalona para evitar el descenso directo en intentar mantener la categoría jugando el playoff de descenso. El Valencia Mestalla llegaba a este". To the right of the article is a Honda advertisement for the "NUEVA HONDA X-ADV POR 125€ AL MES." with details: "Entrada: 2.977€, 36 cuotas de 125€, Cuota final: 5.242,50€, 3,86% TAE." The website's navigation bar includes categories like Fútbol, Motor, Baloncesto, Tenis, Ciclismo, Más deporte, Opinión, AS Color, Resultados, eSports, and a search icon.

Fuente: captura de pantalla de <https://as.com/> informada por el entrevistado

### *Fases del proceso de generación de una información y la interacción del periodista*

El software 'Gabrielle' sigue una serie de pasos para generar las noticias automáticas que se explican a continuación, con el fin de explorar el papel del periodista y la intervención del medio en este proceso.

- **Lectura de datos en tiempo real:** el sistema maneja una serie de datos, como por ejemplo, los resultados de los partidos de fútbol de tercera división.
- **Descarga de los datos e incorporación al sistema:** los datos se incorporan al sistema, previamente entrenado para diferenciar cuáles son relevantes.
- **Maching:** los datos incorporados se organizan gracias a una librería de narrativas automáticas. Estas librerías están formadas por ejemplos o modelos de crónicas, teniendo en cuenta diferentes parámetros, como el nivel de la competición o la tipología de la misma.

El *maching* es el proceso que fusiona los datos extraídos a tiempo real y las narrativas ya existentes en el sistema.

- **Descarga del fichero:** la información se convierte en un fichero para la publicación automática.

Dentro de estas fases de generación de contenido de forma automática, Narrativa posee una amplia librería de narrativas. En este proceso, reciben el asesoramiento de expertos y periodistas. La observación participante realizada en la universidad de Alcalá de Henares, ha permitido explorar cómo comienza el proceso para crear un sistema capaz de generar automáticamente informaciones, esta vez, de carácter político. El próximo reto de Narrativa es automatizar las noticias de las próximas elecciones generales y municipales.

En este caso, el proyecto comenzaba por determinar el universo de narrativas posibles, con el objetivo de concretar qué informaciones podrían automatizarse. Los datos necesarios para realizar las informaciones, provendrían de Indra. A partir de este momento, se realizarían unas estructuras iniciales de posibles casuísticas. Para esta labor, contaban con la colaboración de una periodista encargada de redactar los textos. Del mismo modo, necesitaban de un especialista en electoral para conocer las posibles situaciones o casuísticas, que se estructurarían a través de dominios. En la reunión, se mostraron algunos textos que podían servir para generar el aprendizaje en la máquina. Sin embargo, uno de los aspectos que resaltaron los especialistas fue la necesidad de contar con plantillas o modelos textuales sin interpretación, con una estructura lingüística sin color político ni datos. De esta forma, las plantillas deberían de configurarse de la siguiente forma, que sirve como ejemplo de una posible estructura:

Tabla 1.

*Ejemplo de una estructura de noticia automatizada*

|                                   |   |  |
|-----------------------------------|---|--|
| Primera fase: datos e información | <b>Titular</b>  | El partido X gana las elecciones   |
|                                   | <b>Entradilla basa en datos a tiempo real y datos documentales de anteriores comicios</b> | El partido X ha obtenido X escaños y pasa del X porcentaje al X porcentaje.<br>Sin embargo, el partido X ha obtenido una representación X, a pesar de contar en los anteriores comicios con X porcentajes de voto. |
| Segunda fase: conocimiento        | <b>Cuerpo, riqueza lingüística y conocimiento</b>   | « <b>Tradicionalmente</b> a esta comunidad se suma X», «uno de los pocos <b>bastiones</b> que conserva X», «se trata de un <b>buen dato</b> en la comunidad»   |

Fuente: Elaboración propia a partir de la observación participante llevada a cabo en la universidad de Alcalá de Henares con la startup Narrativa

La complejidad del proceso se centra en una segunda fase descrita de la elaboración de la información, que trata de generar conocimiento y definir términos complejos, como «tradicionalmente», «bastiones» o «buen dato». En este sentido, es importante subrayar la necesidad de seleccionar conceptos objetivos con una definición clara. Por esta razón, las piezas resultantes no contienen una interpretación u opinión en las noticias. Es decir, el proceso

se realiza partiendo de unos datos, que se convierten en información y posteriormente, en un conocimiento más complejo. Quedaría, de esta forma, la generación de informaciones más elaboradas basadas en una interpretación más profunda de los datos o con sesgos políticos deseados, en relación a las líneas editoriales de los medios.

Es, por esta razón, por la que los textos generados a través de técnicas de inteligencia artificial versan sobre campos específicos con un lenguaje previamente definido. Por esta razón, Narrativa genera textos de competiciones deportivas o de asuntos económicos. El lenguaje en estas áreas es más acotado y la disponibilidad de datos abiertos en la red es absoluta.

#### **4.2.3. Creación de infografías automatizadas a través de Prodigioso Volcán.**

En el presente apartado se explicará el proyecto de infografías automatizadas de la empresa Prodigioso Volcán, en colaboración con Software Intelygenz. Esta iniciativa es uno de los tres proyectos en España financiados por Google DNI Found. En este caso, a diferencia de la metodología de Narrativa, se desarrolla una infografía a partir de un texto escrito por un periodista tradicional, relacionada con la pieza en cuestión. Según los expertos de la empresa Software Intelygenz, se trata de un proceso más simple que el de generación del lenguaje, por lo que se puede aplicar a más campos y ámbitos periodísticos. Es decir, se pueden generar infografías automáticas en campos económicos, pero también políticos, demográficos o de otra índole. El primer prototipo creado por Prodigioso Volcán está programado para generar infografías en el ámbito económico, sin embargo, el resultado final pretende abarcar otros campos y ofrecer a los medios de comunicación una herramienta completa para mejorar el entendimiento de sus informaciones.

En primer lugar, es necesario concretar el término «infografía» para conocer qué tipos de resultados se esperan de la aplicación de las técnicas automatizadas. La herramienta es capaz

de generar dos tipos de gráficos con dos diseños diferentes y no es posible generar imágenes más complejas que resuman la información. Estaríamos, por lo tanto, ante los gráficos tradicionales en columnas o líneas. La herramienta resultante es una extensión de Chrome, que automáticamente, genera la infografía gracias al texto escrito por el periodista, que será interpretado a través de técnicas de *machine learning*. A continuación se analiza el proceso de forma más profunda.

### ***Fases del proceso de generación de una información y la interacción del periodista***

Con el objetivo de comprender cómo se desarrolla esta herramienta y la interacción del periodista en la labor de creación de las infografías, se explican en este apartado las diferentes fases del proceso realizadas por la empresa Prodigioso Volcán y por Software Intelligence. Hay que tener en cuenta que la herramienta analizada es el primer prototipo, y por lo tanto, sufrirá modificaciones y adaptaciones en los meses posteriores a la entrega de este Trabajo de Fin de Grado:

- **Redacción de la información principal:** el periodista redacta la información. Existen tres bloques diferenciados en la herramienta de Chrome para facilitar la labor al profesional: un primer bloque para el titular de la noticia, un segundo bloque para la entradilla y un último para el cuerpo de la información.
- **Generación automática del gráfico:** en el prototipo actual, al terminar de redactar la noticia, únicamente es necesario que el periodista haga *click* en la extensión de Chrome para que se genere automáticamente un gráfico relacionado con la información redactada. Las fases previas que permiten actuar a esta herramienta están desarrolladas por Software Intelygenz, y son las siguientes:

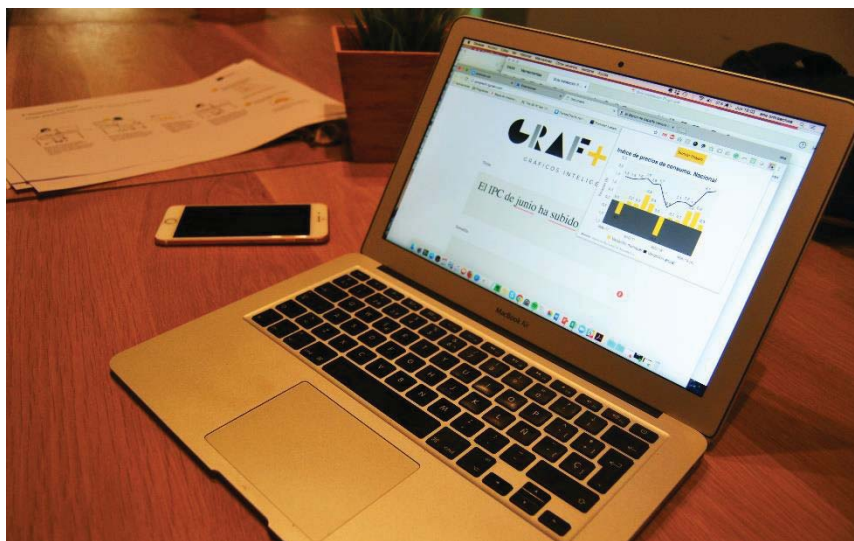


- **Documentación a través de fuentes oficiales:** el programa recaba los datos procedentes de fuentes oficiales, como el INE (Instituto Nacional de Estadística).
- **Entendimiento a través de *matching learning*:** a través de palabras clave y experiencias aprendidas con anterioridad por el programa, se configura un gráfico relacionado con la noticia escrita. Así lo explicaba el encargado del proyecto en Software Intelygenz:

Por ejemplo, si una noticia es sobre el IPC, hay que decir al sistema que versa sobre esta temática. Vamos a coger un número de textos sobre el IPC y decir al sistema «esto es el IPC». El programa va a comenzar a correlacionar, no solo por palabras, si no de forma mucho más profunda, qué el texto versa sobre el IPC. Digamos que a la máquina le damos refuerzos para que pueda aprender y estar segura de que lo hace a través de fuentes fiables.

### Imagen 3

*El prototipo de la extensión de Chrome para la generación automática de infografías*

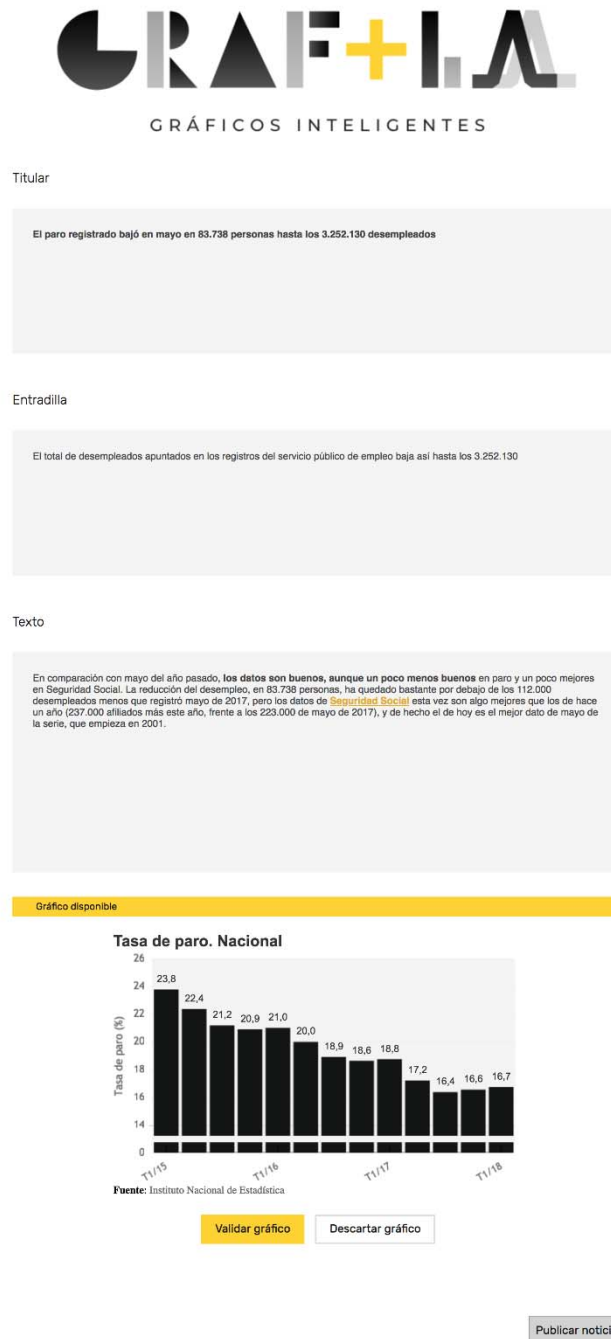


Fuente: Elaboración propia a partir de la información obtenida en Prodigioso Volcán



## Imagen 4

*El prototipo de la extensión de Chrome para la generación automática de infografías*

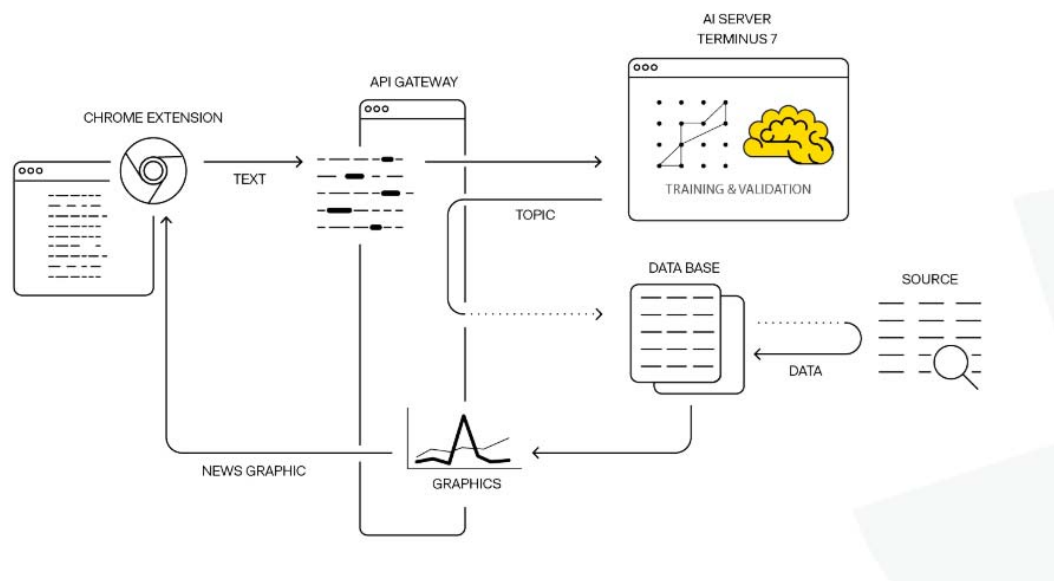


Fuente: Captura de pantalla de la extensión cedida por Prodigioso Volcán

El periodista, por lo tanto, posee un papel fundamental a la hora de redactar la información. Sin embargo, la generación de un gráfico de forma automática le permite prescindir de la necesidad de colaborar con un infografista o realizar el gráfico de forma manual. Así, ambas figuras –periodista e infografista– podrán dedicar su tiempo a tareas más elaboradas. El uso de esta herramienta, además, supone un primer paso para el acercamiento de los medios de comunicación a las técnicas de inteligencia artificial.

## Imagen 5

*Infografía del proceso de generación de gráficos automáticos*



Fuente: Prodigioso Volcán

### 4.3. La relación de los medios de comunicación con las técnicas de inteligencia artificial

Las experiencias explicadas en los apartados anteriores del presente Trabajo de Fin de Grado permiten vislumbrar las principales características de las relaciones existentes en la actualidad entre medios de comunicación y las técnicas de generación automática de contenidos. Los

entrevistados coinciden en señalar que los medios de comunicación se muestran reticentes a la hora de apostar por estas técnicas en sus redacciones. Las experiencias observadas muestran, del mismo modo, como los medios digitales tradicionales son los únicos interesados en adaptarse a la llegada de nuevas tecnologías relacionadas con la inteligencia artificial. Es el caso de los medios de comunicación que publican noticias automatizadas gracias al trabajo realizado por la empresa Narrativa (*El Confidencial, El Independiente, El Español, el Grupo Z, Vocento y MediaPro*). En el caso del proyecto de generación automática de infografías, no es posible establecer unas consideraciones finales a este respecto, puesto que en la fecha de entrega de este Trabajo de Fin de Grado no ha sido expuesto ni presentado a los medios de comunicación.

Por otro lado, es necesario señalar que los medios de comunicación no muestran al lector cuándo una noticia está generada automáticamente. Todos los medios que trabajan con Narrativa publican sus noticias con la firma del periódico digital en cuestión, sin hacer alusión a las técnicas de generación automática ni a la *startup*. Un hecho, que sin embargo, no sucede en otros países en los que trabaja Narrativa, como Alemania o Emiratos Árabes, donde los medios publican la noticia y firman con el nombre de la *startup* y se cita a la fuente de información principal de la que se han extraído los datos.

## **Imagen 6**

*La firma de una información generada automáticamente publicada en el sitio demo de noticias de Narrativa*

Content automatically generated by  Narrativa using data from  BESOCER.

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

#### **4.4. Los límites de la inteligencia artificial: dónde el periodista puede llegar**

Los casos estudiados permiten explorar los límites de la inteligencia artificial en sus aplicaciones al periodismo. En primer lugar, es necesario señalar que nos encontramos ante métodos que funcionan gracias a la disponibilidad de grandes cantidades de datos. Por lo tanto, las informaciones resultantes o los productos finales serán en base a dichos datos. Quedan descartados del ámbito de aplicación todos los géneros periodísticos que necesiten de técnicas de interpretación u opinión, al igual que todas aquellas piezas periodísticas que requieran de una elaboración más compleja. Sin embargo, la existencia de estas técnicas de automatización puede liberar al periodista de los trabajos más rutinarios para poder desempeñar otro tipo de labores. El entrevistado Ramón López de Mántaras, así lo explicaba a través de la investigación sobre abusos a niños por parte de curas en Boston del grupo *Spotlight*:

Este periodismo de investigación nunca se va a sustituir al periodista clarísimamente, porque no tiene muchos elementos rutinarios. Aunque, claro, hay elementos que sí lo son. En *Spotlight* vemos que los periodistas detectan que hay un número anormal de sacerdotes que han tenido bajas por enfermedad o un número anormal de cambios de destinos de los sacerdotes (...) Esto les hizo sospechar que existía algo, que efectivamente era un encubrimiento los más altos responsables de la iglesia de un abuso a niños con estos cambios de destino o bajas. En la película hicieron todo el proceso manualmente, pero ahora, existen proyectos como el ‘proyecto G’<sup>5</sup> que utiliza técnicas de inteligencia artificial para detectar a estos posibles abusadores. Por lo tanto, aquí

---

<sup>5</sup> Con el nombre ‘proyecto G’ bautizan a la herramienta de inteligencia artificial que evalúa los factores de riesgo que sugieren a potenciales violadores, que por su comportamiento pueden ser sospechosos. Información extraída de: <https://www.zeroabuseproject.org/> [Consultado el 27 de junio de 2018]

vemos como herramientas de inteligencia artificial pueden ser de ayuda a los periodistas para detectar estos patrones de comportamiento anormales.

Del mismo modo, es necesario entender que las máquinas no comprenden la información que están procesando, al menos al nivel de un ser humano. Por esta razón, existe un riesgo constante de fallos en las informaciones que deberán ser contrastados por los profesionales del periodismo. Además, cabe señalar que la calidad de los textos que en la actualidad se pueden reproducir a través de técnicas de inteligencia artificial no son de alta calidad, sino que se limitan a reflejar una información basada en datos. Incluso, en el caso de generación de narrativas creativas para otros usos, los resultados son sencillos y se alejan de alcanzar las características de una narrativa compleja.

## 5. Conclusiones

En base a las hipótesis planteadas y los objetivos perseguidos en el presente Trabajo de Fin de Grado, se plantean las siguientes conclusiones y se señalan los límites que derivan de esta investigación y que permanecen abiertos a la investigación académica:

**Primera:** La generación de contenidos periodísticos a través de técnicas de inteligencia artificial es una realidad presente en nuestro país, ya que se están publicando noticias generadas automáticamente en medios digitales nacionales. Esto se ve reflejado en las noticias publicadas por diarios como *El Confidencial*, *El Independiente*, *El Español*, *El As* o *Sport* gracias al método de generación de noticias de la *startup* Narrativa. La incorporación de estas nuevas técnicas proviene de empresas o *startups* especializadas en inteligencia artificial, como es el caso de Narrativa o Software Intelygenz. Por lo tanto, en los medios de comunicación únicamente se realiza una labor de implementación de las mismas, sin desarrollar en laboratorios u otro tipo de equipos de trabajo estas técnicas.

Del mismo modo, las empresas consultadas inciden en el creciente interés de los medios de comunicación en las técnicas de automatización en los últimos dos años. Sin embargo, también señalan la reticencia de los medios más tradicionales a incorporarlas en sus redacciones. Los medios de comunicación digitales ocuparían, de esta forma, un papel relevante en la incorporación de la inteligencia artificial en las dinámicas periodísticas. Hasta el momento, y según la muestra explorada en el presente trabajo, existen pocas empresas que apuesten por la generación automática de contenidos. Según las fuentes consultadas, únicamente Narrativa genera textos automatizados para medios de comunicación en nuestro país.

Así, la primera hipótesis del presente trabajo referida a un aumento de la relevancia del periodismo automatizado panorama nacional a través de *startups* que apuestan por ofrecer noticias u otro tipo de contenidos generados automáticamente a los medios de comunicación tradicionales, se confirma. Del mismo modo, habría que matizar que el trabajo en las técnicas de inteligencia artificial también se desarrolla en empresas con una larga trayectoria como Software Intelygenz. Aunque, sí es cierto que los proyectos se engloban siempre en iniciativas de innovación lejanas a las redacciones o a los conglomerados mediáticos. La publicación de 12.000 noticias semanales generadas automáticamente a través del software ‘Gabrielle’ de Narrativa demuestra el creciente interés en España por el uso de material automatizado.

**Segunda:** Los contenidos generados automáticamente se basan en el aprendizaje computacional o *maching learning*, a través de la ingesta de grandes cantidades de datos. En los dos casos estudiados, la generación de contenido era posible gracias a una gran cantidad de datos existentes, que sirven de base para la redacción o la creación de contenido. Además de estos datos disponibles a tiempo real o a través de técnicas de documentación, los sistemas necesitan de estructuras o ejemplos para poder generar un contenido similar. En el caso de Narrativa, las noticias resultantes tienen como base principal la ingesta de datos y el aprendizaje

a través de estructuras narrativas previamente generadas. Lo mismo sucede en el caso de la generación automática de infografías de Prodigioso Volcán e Intelygenz, se trata de un sistema que aprende a través de ejemplos para reproducir un determinado comportamiento. De esta forma, las informaciones resultantes en texto (generadas por la empresa Narrativa) únicamente abarcan dos campos periodísticos, el deporte y la economía. Por otro lado, la generación automática de contenidos visuales o infografías, cubre un mayor número de campos, aunque en la entrega de este Trabajo de Fin de Grado el prototipo generado por la empresa Prodigioso Volcán únicamente realiza infografías económicas. Sin embargo, en ambos casos la información tiene que estar sustentada en datos para que la máquina pueda generar contenido.

De esta forma, las segunda de las hipótesis planteadas referida a la mayor incidencia del periodismo robot en las informaciones basadas en datos o estadísticas se confirma. Del mismo modo, por el momento, las aplicaciones reales de la inteligencia artificial en la muestra estudiada no han conseguido modificar otras informaciones generadas a través de otras técnicas no relacionadas con el *big data*. Por lo tanto, quedarían campos por explorar y dinámicas periodísticas irremplazables, necesarias de una mayor elaboración y una dificultad que traspasa el entendimiento de una máquina.

**Tercera:** Las informaciones generadas a través del software ‘Gabrielle’ no necesitan de la intervención del periodista en la redacción de la información ni en la publicación. Por otro lado, la generación automática de infografías necesita de una pieza periodística previa, realizada por un profesional, para generar automáticamente una infografía. Sin embargo, si aunásemos los dos procesos, sería posible generar una noticia económica a través de técnicas de inteligencia artificial e insertar en la misma una infografía generada automáticamente, sin la necesidad de contar con un profesional.

Por lo tanto, la tercera de las hipótesis del presente trabajo, referida a la imprescindible intervención del periodista en la generación de informaciones, a pesar de utilizar técnicas de inteligencia artificial, se refuta. Los dos casos estudiados demuestran que en la actualidad ya es posible generar automáticamente un tipo concreto de informaciones basadas en datos a través de técnicas de aprendizaje.

**Cuarta:** Los medios de comunicación tradicionales se muestran reticentes a incorporar nuevos modelos de generación automática de contenido, aunque en los últimos dos años, la tendencia ha sido creciente en los medios digitales. Sin embargo, cabe destacar que no hacen pública la utilización de estas técnicas en sus informaciones, puesto que en todos los casos estudiados, el medio de comunicación firma con el su propio nombre y no comunica al lector la fuente de información. Del mismo modo, los entrevistados coinciden en señalar a la empresa periodística como un sector difícil para implementar técnicas de inteligencia artificial, a diferencia de otros más accesibles, como el financiero. Por otro lado, la experiencia de la empresa Narrativa con medios de comunicación de otros países sitúa a España en un lugar inferior en el uso de estas técnicas, razón por la que Narrativa explora nuevos sectores de trabajo.

Así, la última de las hipótesis planteadas, referida a la reticencia de los medios de comunicación a incorporar nuevos modelos de generación automática de contenido por temor a una desestructuración de las dinámicas tradicionales y pérdida de puestos de trabajo queda confirmada, a pesar del uso creciente de estas técnicas por parte de diarios digitales españoles.

En definitiva, la presente investigación demuestra que el periodismo robot es una realidad presente en España, a pesar de la reticencia de los medios más tradicionales a implementarlo en sus redacciones. De esta forma, se demuestra que las tareas más rutinarias realizadas a lo largo de la historia por periodistas, pueden ser sustituidas por un sistema. Sobre todo, en las



áreas donde existe un gran número de datos disponibles, como son los deportes o la economía y finanzas. Además, cabe destacar que los proyectos estudiados demuestran la existencia de empresas o *startups* en España interesadas en la innovación en el campo periodístico.

Este Trabajo de Fin de Grado (TFG) propone estas conclusiones, aplicables a los casos estudiados. Sin embargo, cabe destacar que al tratarse de un campo poco estudiado en nuestro país, existen dificultades a la hora de extrapolar los datos al universo estudiado. Sería interesante, además, conocer la opinión de los medios de comunicación en cuestión y estudiar en qué aspectos influyen estas dinámicas en las redacciones. Por esta razón, el presente Trabajo de Fin de Grado puede definirse como el comienzo de un estudio más profundo de las técnicas de inteligencia artificial, con el objetivo de conocer sus repercusiones en las dinámicas periodísticas.

## 6. Bibliografía

- Arcila-Calderón, C, Barbosa-Caro, E y Cabezuelo-Lorenzo, F. (2016). Técnicas *Big Data*: Análisis de textos a gran escala para la investigación científica y periodística. *Revista internacional de Información y Comunicación*, 25 (4). Recuperado de:  
<http://www.elprofesionaldelainformacion.com/contenidos/2016/jul/12.html>
- Arjen van Dalen. (2012). The Algorithms behind the headlines. *Journalism Practice*, 6 (5-6)  
Recuperado de:  
<https://www.tandfonline.com/doi/abs/10.1080/17512786.2012.667268>
- Astrid Gynnild (2014). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism*, 15 (6). Recuperado de:  
<http://journals.sagepub.com/doi/abs/10.1177/1464884913486393>
- Bercovici, J. (2010, 11, 29). Can you tell a robot write this? *Forbes*. Recuperado de:  
<https://www.forbes.com/sites/jeffbercovici/2010/11/29/can-you-tell-a-robot-wrote-this-does-it-matter/#77f1ec02565b>
- Bernard, R. and Gravlee, C. (2014). *Handbook of Methods in Cultural Anthropology*. Plymouth, United Kingdom: AltaMira Press
- Broussard, M. (2014): Artificial Intelligence for Investigative Reporting. *Digital Journalism*, 3 (6). Recuperado de:  
<https://www.tandfonline.com/doi/full/10.1080/21670811.2014.985497>

Bryman, A. (1984). The Debate about Quantitative and Qualitative Research: A Question of Method or Epistemology? *The British Journal of Sociology*, 35 (1), 75-92.

Recuperado de: <http://www.jstor.org/stable/590553>

Bunz, M. (2015). In the US, algorithms are already reporting the news. *The Guardian*.

Recuperado de: <https://www.theguardian.com/media/pda/2010/mar/30/digital-media-algorithms-reporting-journalism>

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big data and Society*, 3 (1). Recuperado de:

<http://journals.sagepub.com/doi/abs/10.1177/2053951715622512>

Callejo, J (2002). Observación, entrevista y grupo de discusión. El silencio de tres prácticas de investigación. *Revista Española de Salud Pública*, 76 (5), 409-4122. Recuperado

de: <https://www.scielosp.org/article/resp/2002.v76n5/409-422/es/>

Carlson, M. (2014). The Robotic Reporter, Digital Journalism. Recuperado de: <https://scihub.tw/http://www.tandfonline.com/doi/abs/10.1080/21670811.2014.976412>

Cervera, J. (2017). El futuro del periodismo es Ciborg. *Cuadernos de periodistas* (34).

Recuperado de:

[http://www.cuadernosdeperiodistas.com/media/2017/11/Jose\\_cervera.pdf](http://www.cuadernosdeperiodistas.com/media/2017/11/Jose_cervera.pdf)

Crerwall, (2014). Enter the robot Journalist. *Journalism Practise*, 8 (5), 519-531. Recuperado

de: <https://www.tandfonline.com/doi/abs/10.1080/17512786.2014.883116>

Dhar, V. (2012). Data Science and Prediction. *Communications of the ACM*, 56 (12), 64-73.

Recuperado de: <https://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/abstract>

Ford, M. (2013). Could artificial intelligence create an unemployment crisis?

*Communications of the ACM*, 56 (7), 37-39. Recuperado de:

<https://cacm.acm.org/magazines/2013/7/165475-could-artificial-intelligence-create-an-unemployment-crisis/abstract>

Janesick, V. J. (1994). *The dance of qualitative research design: Metaphor, methodolaty, and meaning*. Recuperado de:

[https://www.researchgate.net/publication/232497144\\_The\\_dance\\_of\\_qualitative\\_research\\_design\\_Metaphor\\_methodolaty\\_and\\_meaning](https://www.researchgate.net/publication/232497144_The_dance_of_qualitative_research_design_Metaphor_methodolaty_and_meaning)

Kawulich, B. (2005). La observación participante como método de recolección de datos.

*Forum: Qualitative Social Research*, 6 (2). Recuperado de:

<http://diverrisa.es/uploads/documentos/la-observacion-participante.pdf>

Konstantin Nicholas Dörr, K. & Hollnbuchner, K (2016). Ethical Challenges of Algorithmic

Journalism, Digital Journalism. *Digital Journalism*, 5 (4), 404-4019. Recuperado de:

<https://www.tandfonline.com/doi/abs/10.1080/21670811.2016.1167612>

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási A.L., Brewer, D., Christakis, N.,

Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. y

Alstyne. N. Computational Social Science. *Science*, 323 (5915), 721-723. Recuperado

de: <http://science.sciencemag.org/content/323/5915/721.long>

Lewis, S y Westlund (2014). Big Data and Journalism. *Digital Journalism*, 3 (3), 447-466.

Recuperado de: <https://apps.cla.umn.edu/directory/items/publication/324155.pdf>

López, R. E. y Deslauriers, J. (2011). La entrevista cualitativa como técnica para la investigación en Trabajo Social. *Margen* (61). Recuperado de:

<https://dialnet.unirioja.es/servlet/articulo?codigo=3756178>

Mejía, J. (2000). El muestreo en la investigación cualitativa. *Investigaciones sociales* (5).

Recuperado de:

<http://revistasinvestigacion.unmsm.edu.pe/index.php/sociales/article/viewFile/6851/6062>

Noam Latar (2015). The Robot Journalist in the Age of Social Physics: The End of Human Journalism? *The New World of Transitioned Media*, 65-80. Recuperado de:

[https://link.springer.com/chapter/10.1007/978-3-319-09009-2\\_6#citeas](https://link.springer.com/chapter/10.1007/978-3-319-09009-2_6#citeas)

Salaverría, R. (2014). Tendencias del periodismo mundial. Periodismo en 2014: balance y tendencias. *Cuadernos de periodistas*, (29). Recuperado de:

<https://dialnet.unirioja.es/servlet/articulo?codigo=4922109>

Salaverría, R. y García, J.A. (2008). La convergencia tecnológica en los medios de comunicación: retos para el periodismo. *Trípodos*, (23). Recuperado de:

<https://www.raco.cat/index.php/Tripodos/article/viewFile/118910/154114>

Sánchez, H y Sánchez, M. (2017). Los bots como servicio de noticias y de conectividad emocional con las audiencias. El caso de *Polibot*. *Doxa.Comunicación*, (25), 63-84.

Recuperado de:

Aguado Carro, Paloma (2018): *La inteligencia artificial aplicada al periodismo: infografías automatizadas y narrativa robot*. Trabajo de Fin de Grado de Periodismo. **Universidad de Valladolid**

[http://dspace.ceu.es/bitstream/10637/8765/1/Bots\\_HadaSanchez\\_MariaSanchez\\_Doxa\\_2017.pdf](http://dspace.ceu.es/bitstream/10637/8765/1/Bots_HadaSanchez_MariaSanchez_Doxa_2017.pdf)

Terminus7 y Intelygenz Academy Team (20128). *Artificial Intelligence: understanding machine learning*.

Valero-Pastor, J. M. y González-Alba, J. A. (2018): Las startups periodísticas como ejemplos de innovación en el mercado mediático español. Estudio de casos. *Revista Latina de Comunicación Social* 73 (4), 556-582. Recuperado de:  
<https://dialnet.unirioja.es/servlet/articulo?codigo=6341534>

## 7. Anexos

### 7.1. Ficha a rellenar durante la entrevista

|                                      |  |                                   |  |
|--------------------------------------|--|-----------------------------------|--|
| <b>Entrevistado:</b>                 |  |                                   |  |
| <b>Cargo y funciones:</b>            |  |                                   |  |
| <b>Fecha de entrevista:</b>          |  |                                   |  |
| <b>Lugar de la entrevista:</b>       |  |                                   |  |
| <b>Hora de inicio de entrevista:</b> |  | <b>Hora de fin de entrevista:</b> |  |
| <b>Nombre de entrevistador (es):</b> |  |                                   |  |

El entrevistado autoriza la grabación (audio) del contenido de la entrevista únicamente con fines investigadores en el marco del proyecto «La inteligencia artificial aplicada al periodismo: infografías automatizadas y narrativa robot», enmarcado en los Trabajos de Fin de Grado de Periodismo de la Universidad de Valladolid en el curso 2017-2018.

Sí

No

**Comentarios (si fuese oportuno) del entrevistador(es):**

## 7.2. Bloques de contenidos de las entrevistas

| Bloques   | Objetivos   |
|---|---|
| <b>1. Situación actual de las investigaciones relacionadas con la inteligencia artificial en España</b> | Conocer cuál es la situación actual de los métodos de inteligencia artificial en España aplicados a diferentes áreas, y cómo se están desarrollando las investigaciones en este campo. Pretende así, ser un contexto general para la investigación.   |
| <b>2. La inteligencia artificial aplicada al periodismo: su actualidad</b>                              | Determinar cómo se pueden aplicar las técnicas de inteligencia artificial al periodismo, y más concretamente, a los medios de comunicación. Se pretende de esta forma, realizar una radiografía del estado actual en el que se encuentra la aplicación de métodos automatizados en las redacciones de los medios. |
| <b>3. La inteligencia artificial aplicada al periodismo: el futuro en España</b>                        | Analizar el futuro de las técnicas de inteligencia artificial. Así, se podrán vislumbrar los posibles cambios que se pueden generar en los próximos años.   |



### 7.3. Preguntas de las entrevistas

#### **Bloque 1: Situación actual de las investigaciones relacionadas con la inteligencia artificial en España.**

- ¿Cómo definiría la inteligencia artificial, el *machine learning* y el *big data* y el *data mining*?
- ¿En qué situación se encuentra el uso de técnicas relacionadas con la Inteligencia Artificial en los diferentes sectores? En comparación con otros países, ¿cómo ha evolucionado España en este campo?
- ¿En qué proyectos trabaja en la actualidad el Instituto de Investigación de Inteligencia Artificial/el grupo de trabajo de Generación de Narrativas? ¿Por qué se centra en estos campos? ¿Existe algún proyecto relacionado con el periodismo, los medios de comunicación, el marketing o la escritura creativa en el Instituto de Investigación de Inteligencia Artificial/en el grupo de trabajo? En caso afirmativo, ¿cómo se ha desarrollado este proyecto y en qué consistía?

#### **Bloque 2: La inteligencia artificial aplicada al periodismo: su actualidad**

- ¿En qué momento se comienza hablar en España de unas implicaciones reales de los métodos de automatización e inteligencia artificial en el periodismo?
- Hay proyectos que demuestran que, a través de una lectura de grandes cantidades de datos, pueden generarse narrativas automáticas similares a las realizadas por un periodista, ¿cómo es el proceso para llevar a cabo estas dinámicas?

- En Estados Unidos, agencias de noticias como *Associated Press*, utilizan estas técnicas para determinadas informaciones, como pueden ser las deportivas, aunque también sobre economía ¿qué otras tareas podría realizar un robot en las redacciones?
- Estas informaciones, ¿serían la base para una posterior redacción más profundizada, o por el contrario, podrían construirse relatos informativos lo suficientemente bien elaborados para un consumo directo? ¿podrían construirse relatos, más allá de informativos, con características como la creatividad?
- Los medios de comunicación también utilizan *bots* para distribuir noticias e informaciones con el objetivo de facilitar una información personalizada a cada usuario, ¿cómo pueden llevarse a cabo este tipo de trabajos y qué coste supondría para una empresa del sector periodístico?
- La robotización también alcanza otros niveles, como puede ser la comunicación de los partidos políticos o la comunicación institucional. ¿En qué grado se está haciendo uso de herramientas automatizadas en estas áreas?
- ¿En qué lugares se está experimentando con iniciativas de inteligencia artificial aplicadas al periodismo (es decir, en el ámbito académico, empresarial...)? ¿Se podría integrar este proceso en las propias redacciones de los medios o en los laboratorios de innovación, en vez de contratar empresas externas?

### **Bloque 3: La inteligencia artificial aplicada al periodismo: el futuro en España**

- ¿Cuáles son los retos a los que se ven expuestos los periodistas con la implantación de métodos automatizados en las redacciones? ¿Estaremos ante una nueva forma de hacer periodismo y unos periodistas con otro tipo de formación académica?

- Uno de los mayores temores de los investigadores es que los robots acaben con puestos de trabajo, es decir, que desempeñen tareas que tradicionalmente correspondían al periodista con un coste más bajo. ¿Hasta qué punto esta afirmación podría ser una realidad presente?
- ¿Cuándo estarán presentes los robots en las redacciones de los medios de comunicación en España? ¿Existirá una colaboración necesaria entre empresas que desarrollan estas tecnologías y los directores de los medios de comunicación?