**TABLE OF CONTENTS**

[1 INTRODUCTION 3](#_Toc135757411)

[2 OBJECTIVE AND RELEVANCE 4](#_Toc135757412)

[3 METHODOLOGY 5](#_Toc135757413)

[3. 1 Data sources 5](#_Toc135757414)

[4 ANALYSIS 5](#_Toc135757415)

[4.1 Data extraction 5](#_Toc135757416)

[4.2 Data cleaning 5](#_Toc135757417)

[4.3 Text analysis 5](#_Toc135757418)

[4.4 Data visualization 5](#_Toc135757419)

[5 RESULTS 5](#_Toc135757420)

[6 CONCLUSIONS AND LIMITATIONS 5](#_Toc135757421)

[7 REFERENCES 5](#_Toc135757422)

**Abstract**

# INTRODUCTION

En la última década en España, al igual que en el resto del mundo occidental, se ha experimentado un creciente interés por el desarrollo de la Inteligencia Artificial. Este interés ha venido precedido por el lanzamiento de diversas herramientas que incorporan esta tecnología. Ejemplo de ello son los asistentes virtuales inteligentes como Siri de Apple, Google Assistant, Alexa de Amazon o Microsoft Cortana. También empresas como Tesla han avanzado mucho en el desarrollo de vehículos autónomos que incorporan la IA para reconocer el entorno e incluso tomar decisiones autónomas. Otros campos en los que esta tecnología ha supuesto una revolución son los sistemas de recomendación en diversas plataformas de entretenimiento (por ejemplo, Netflix o Spotify); en la medicina, donde se han desarrollado algoritmos para el diagnóstico médico; y en el procesamiento del lenguaje natural que ha permitido el desarrollo de chatbots como Chat-GPT o Midjourney.

La irrupción de la IA en la sociedad ha puesto sobre la mesa diversos desafíos éticos y sociales derivados de su uso. Cada vez más gente muestra interés por esta tecnología y se pregunta cosas sobre ella. Los medios de comunicación han plasmado este debate a la vez que han sido líderes de opinión mediante la labor divulgativa de algunos temas que todavía no habían llegado a la sociedad. Además, en los últimos años hemos visto como esta cuestión ha llegado cada vez más lejos, haciendo que también los representantes políticos se vean en la necesidad de sentarse a tomar decisiones sobre la regulación de la IA. En el contexto de la Unión Europea se está debatiendo la propuesta de la Ley de Inteligencia Artificial (cita) que tiene entre otros propósitos el de establecer un sistema para evaluar el riesgo que una tecnología que incorpore la IA puede suponer para la salud y la seguridad de las personas (Newtral).

Este Trabajo de Final de Master se adentra en el debate a través de las noticias sobre Inteligencia Artificial publicadas en dos medios de comunicación digitales españoles. A través de técnicas de escrapeo web se obtendrán los textos de las noticias que posteriormente se analizarán mediante análisis de texto. Este trabajo pretende contribuir de forma empírica al análisis de noticias, es por ello que todo el código usado para realizar el proceso de extracción de datos, limpieza, análisis y visualización quedará adecuadamente recopilado en varios archivos formato rmd que estarán subidos en el siguiente enlace de GitHub. El trabajo se estructura en seis apartados diferentes: primero se explican los objetivos y la relevancia de este, haciendo un pequeño análisis de la literatura existente sobre el tema. Posteriormente se presentará de forma breve la metodología y las fuentes de datos empleadas. En el apartado principal del trabajo se explica cómo se ha realizado la extracción de datos, su limpieza, el análisis de texto y la visualización. Finalmente, se exponen los resultados y se esbozaran las conclusiones.

# OBJECTIVE AND RELEVANCE

El objetivo de este trabajo es realizar una aportación empírica en el ámbito de la extracción de datos de noticias de periódicos digitales y su análisis de texto a través del entorno de programación R. El código creado para realizar la extracción del contenido de las noticias será replicable, por lo que cualquier persona que quiera realizar un análisis similar podrá hacer uso de este código y adaptarlo para extraer la selección de noticias que necesite. Esto es importante porque actualmente estos periódicos no tienen sus noticias en ninguna API ni en otro formato accesible, por lo que la extracción de datos mediante la técnica de web scraping es la única forma posible de analizar el contenido. Por otra parte, el análisis de texto a través de R también facilita mucho poder extraer conclusiones de grandes volúmenes de texto sin tener que recurrir a su lectura pormenorizada, lo cual supondría invertir mucho tiempo. R cuenta con una gran cantidad de paquetes especializados en el procesamiento del lenguaje natural (NLP), que permiten procesar y analizar los textos de manera más rigurosa y eficiente, así como herramientas de visualización de datos textuales que ayudan a comunicar estos resultados de forma comprensible. Este código también estará disponible para quien quiera realizar el análisis de otra base de datos de texto diferente a la que se va a usar en este trabajo.

El análisis de las noticias publicadas en los medios de comunicación es útil especialmente cuando no se dispone de otra fuente de datos sobre la opinión de los ciudadanos en torno a un tema concreto. Como la irrupción de la Inteligencia Artificial es relativamente reciente, no hay apenas encuestas ni material cualitativo (entrevistas, grupos de discusión…) que permita conocer qué piensa la ciudadanía de esta tecnología. En este contexto, analizar la prensa permite obtener una información muy valiosa sobre cuáles son los temas principales de los que se habla, qué tipo de datos se divulgan y cuáles son los temas conflictivos que se están poniendo sobre la mesa. Además, este trabajo va a incluir noticias de periódicos de diferentes ideologías, lo que también permite comparar el posicionamiento de cada uno de ellos respecto de la inteligencia artificial. En definitiva, el análisis del texto de las noticias sobre inteligencia artificial que se va a realizar en este trabajo va a permitir identificar tendencias y temas principales, evaluar el tono y la opinión predominante, detectar sesgos y analizar la cobertura mediática de este tema en el periodo de tiempo analizado.

Existen ejemplos de otros trabajos de investigación que también han aplicado técnicas de análisis de texto para estudiar el contenido de las noticias de medios de comunicación. Añadir papers.

El trabajo de investigación se estructura en cuatro apartados diferentes. En primer lugar se va a explicar la metodología empleada y las fuentes de donde se ha extraído la información. Este apartado es especialmente importante puesto que al tratarse de un trabajo empírico tiene una especial relevancia el trabajo de extracción de datos realizado. Por ello, se dedicará un apartado propio a explicar cómo se ha llevado a cabo el proceso de extracción de datos mediante el lenguaje de programación R y la herramienta RStudio. Posteriormente, se explicará cómo se han limpiado los datos. El cuarto apartado consistirá en un análisis de texto donde también expondrán las visualizaciones realizadas que permitirán profundizar en ese análisis. El trabajo finaliza con un breve apartado con las conclusiones y las limitaciones del mismo.

# METHODOLOGY

## 3. 1 Data sources

En este trabajo de investigación se usarán datos extraídos de dos fuentes primarias, la web de dos periódicos digitales españoles: elDiario.es (<https://www.eldiario.es/>) y El Mundo ([elmundo.es](https://www.elmundo.es/)). Se ha escogido estos dos periódicos porque se trata de dos periódicos generalistas de ámbito nacional, por tanto, son comparables entre sí. Además, tienen una línea editorial ideológicamente diferente, por lo que la comparación puede resultar más interesante. Mientras que elDiario.es se sitúa a la izquierda en el espectro ideológico, El Mundo se sitúa a la derecha.

## 3. 2 Sample

## 3. 3 Data extraction

+1000

## 3. 4 Data cleaning

# ANALYSIS

## 4.3 Text analysis

## 4.4 Data visualization

# RESULTS

+1000

# CONCLUSIONS AND LIMITATIONS

+500

# REFERENCES

* [**https://www.newtral.es/ley-inteligencia-artificial-ia-chatgpt-union-europea-ue/20230427/**](https://www.newtral.es/ley-inteligencia-artificial-ia-chatgpt-union-europea-ue/20230427/)