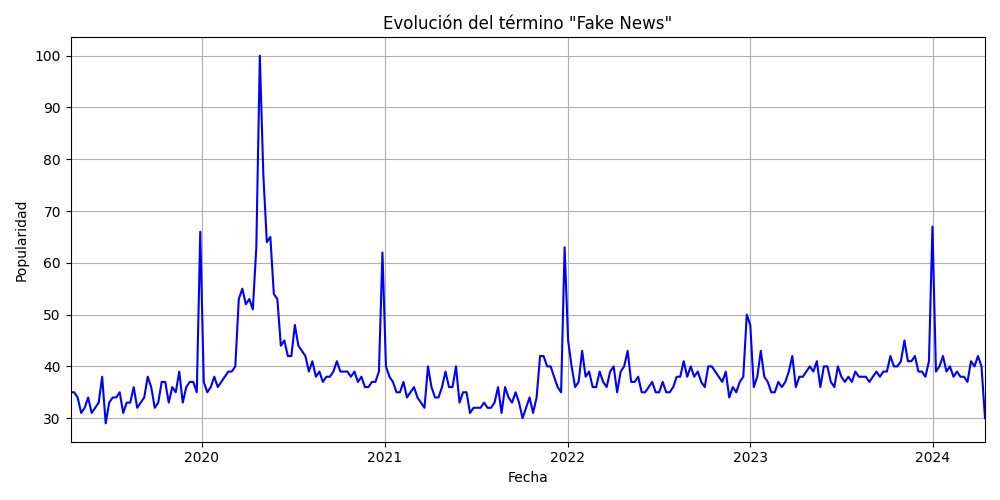
Capítulo 1

## **Introducción**

En plena era digital, el fenómeno de las noticias falsas, o “*fake news”,* ha emergido como una poderosa fuerza que trasciende fronteras lingüísticas, influenciando de manera significativa la percepción pública, la toma de decisiones o la estabilidad democrática en todo el mundo. Este impacto se ve aumentado por la rapidez con la que la desinformación se propaga a través de las redes sociales y otros canales de información.

Uno de los principales objetivos de las *fake news* es la manipulación de la opinión de la sociedad para fines políticos o sociales. Actualmente, podemos apreciarlo con claridad con el conflicto entre Israel y Palestina, donde a diferencia de otros conflictos bélicos en el pasado las nuevas armas toman forma de ‘post’ o publicación cuya crudeza interpela directamente a los sentimientos de los usuarios. [1]



**Fig 1 - Evolución del término "Fake News" durante los años**

En la siguiente gráfica podemos observar el crecimiento de la popularidad del término *fake news,* destacando 3 escenarios históricos relevantes. El primero es en plena pandemia y se extiende hasta finales de 2021. El segundo en 2022, el conflicto entre Ucrania y Rusia. Por último, a finales de 2023, cuando el conflicto entre Israel y Palestina estaba en su punto álgido.

Aunque a día de hoy la mayoría de *fake news* que vemos provienen de usuarios anónimos con un bajo nivel cultural, detrás de la propagación de bulos y desinformación se encuentran partidos políticos o grandes empresas.

Un ejemplo claro fue el caso de *Cambridge Analytica*, una empresa que saltó a la fama por su supuesta participación en la campaña electoral presidencial de Estados Unidos en 2016 o con el Brexit en Europa. La propia empresa tiene en la portada de su página web el slogan “Data drives all we do. Cambridge Analytica uses data to change audience behaviour” [2]

En el siguiente artículo del periódico EL PAÍS se hace especial énfasis en la comunidad hispanohablante en Estados Unidos señalando que esta población es la más afectada por la desinformación en el país. Las traducciones al español pueden transformar un contexto en otro totalmente diferente, algo de lo que se aprovechan los medios de comunicación en la zona. Temas muy críticos como la salud pública o los derechos de inmigración son los que más preocupan a los habitantes hispanohablantes. [3]

Con estos ejemplos de la actualidad social, subrayan la importancia para los ciudadanos de saber diferenciar entre información verídica y falsa, ya que como se ha visto tiene un gran impacto tanto en la sociedad como en el individuo.

## **1.1 Motivación**

El desafío de diferenciar entre noticias falsas y verdaderas se ve a día de hoy obstaculizado por la falta de tiempo en la actualidad. Cada día recibimos más información de diferentes medios como redes sociales, medios tradicionales entre otros. Las personas no disponen del tiempo que se necesita para contrastar que la noticia que estamos leyendo es falsa o no.

Esta situación se agrava más si hablamos de las noticias en español. La comunidad hispanohablante se enfrenta a barreras adicionales en el acceso a información confiable en su propio idioma debido a que el idioma que predomina en estos casos es el inglés.

La creación de un sistema automático de detección de *fake news* que se enfoque en el español se vuelve crucial. El objetivo es abordar el problema con el texto como elemento principal para la decisión de la veracidad de la información. Se realizará desde el enfoque del Procesamiento del Lenguaje Natural (PLN) para analizar y evaluar la veracidad del texto, y no utilizando otros métodos como las fuentes o la red subyacente de comunicación entre usuarios que comparten la noticia falsa.

## **1.2 Objetivos**

El objetivo principal de este trabajo fin de grado es el desarrollo de un sistema de detección de desinformación basado en RoBERTa.

A continuación, se enumeran los objetivos específicos necesarios para la realización del objetivo principal:

1. **Analizar el estado del arte del problema**: Se estudiará el estado del arte del problema de detección de *fake news, Deep Learning* y PLN.
2. **Analizar los conjuntos de datos disponibles**: Se analizarán los conjuntos de datos más populares y recientes en la literatura.
3. **Desarrollar un sistema de *Deep Learning* para resolver el problema**: Se desarrollará un sistema de *Deep Learning* para resolver el problema utilizando únicamente el texto de la noticia (título y cuerpo), sin ningún tipo de información adicional.
4. **Implementar una interfaz web para el problema**: Se implementará una interfaz web para mostrar el funcionamiento del sistema de forma interactiva.

## **1.3 Planificación**

## **1.4 Estructura**

Capítulo 2

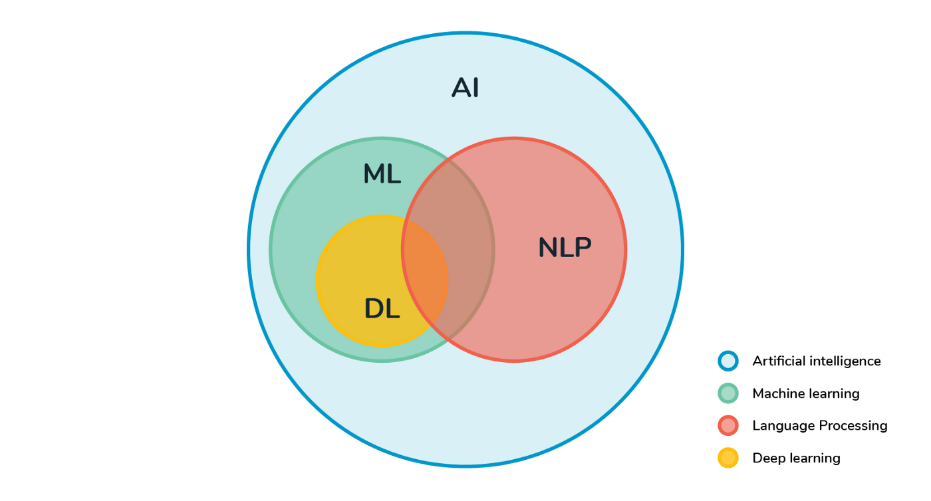
## **Estudio de Fake News y Procesamiento del Lenguaje Natural**

En este capítulo, se establece el marco teórico del trabajo de investigación, enfocado en el estudio y comprensión de las *fake news,* Procesamiento del Lenguaje Natural, *Transformers,* BERT y RoBERTa.

## **2.1 Fake News**

## **2.2 Procesamiento del Lenguaje Natural**

El procesamiento del lenguaje natural es un subcampo de la lingüística, ciencia de computadores e inteligencia artificial dedicado a diseñar métodos y algoritmos que toman como entrada o salida lenguaje natural no estructurado. El objetivo es que un ordenador sea capaz de generar y “entender” los contenidos de un texto para así poder interactuar con él.



**Fig 2 - Diagrama NLP**

En los últimos años, el procesamiento de grandes cantidades de información a través de texto está cada vez más presente, surgiendo multitudes de dificultades para poder tratarlos. Es por ello que esta disciplina está alcanzando un importante papel en la actualidad. Algunas de las tareas más típicas que aborda el procesamiento del lenguaje natural son:

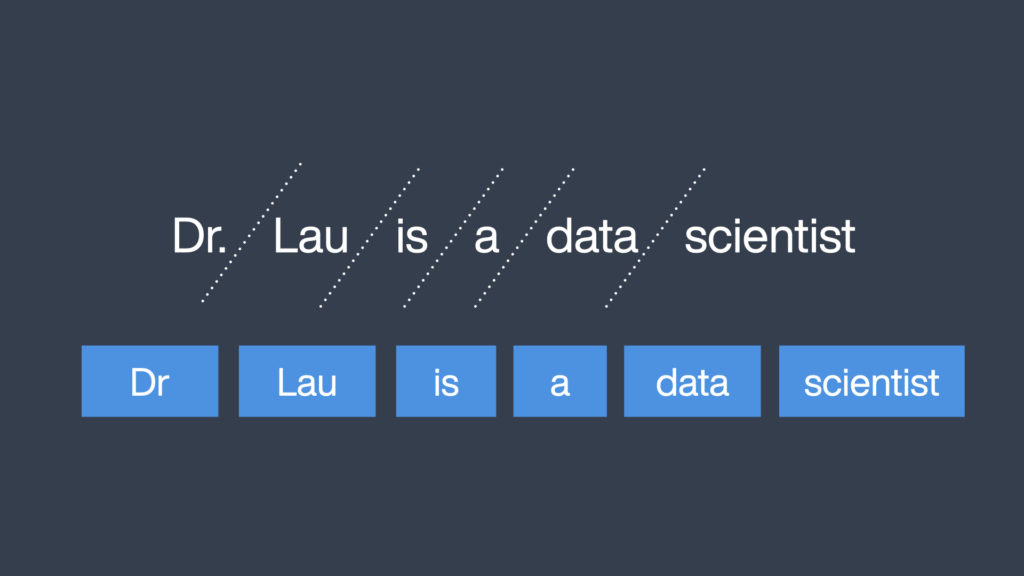
* **Reconocimiento de voz**. Dado un audio de una persona hablando, convertirlo en texto.
* ***Text-to-speech****.* Al contrario que el reconocimiento de voz, transforma el texto en audio.
* **Clasificación de texto.** Asignar a un texto o un documento una o más categorías.
* **Traducción automática.** Traducción automática de texto de un idioma a otro.
* **Generación automática de texto.** Dadas unas características generar un texto que las cumpla.
* **Análisis de sentimientos.** Caso específico de la clasificación de texto en el que se le asigna un sentimiento al texto.ç

Cuando tenemos un problema de PLN, podemos distinguir dos aspectos esenciales:

* La parte “**lingüística**”, que consiste en preprocesar y transformar la información de entrada en una serie de datos explotables.
* La parte de “**aprendizaje automático**”, que se basa en la aplicación de modelos de *Machine Learning* o *Deep Learning* a ese conjunto de datos.

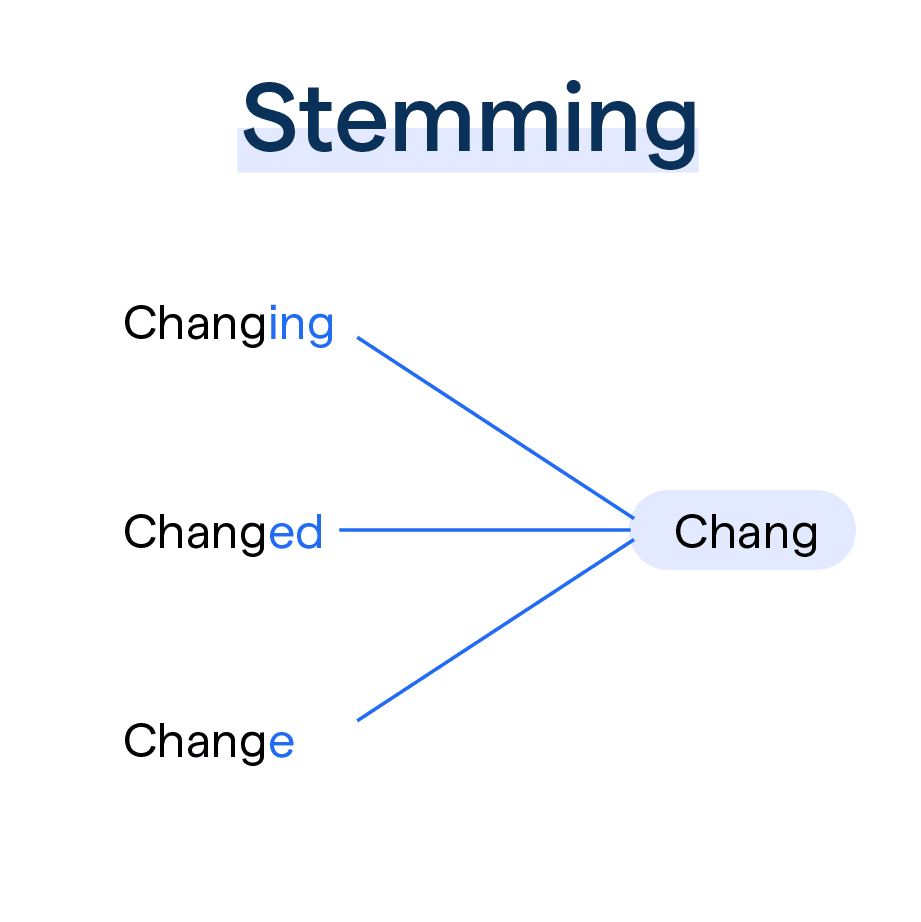
La fase de preprocesamiento consiste en limpiar y preparar el texto para que un algoritmo PLN pueda analizarlo. Entre las principales etapas que se siguen, encontramos:

* **Limpieza del texto**: eliminar signos de puntuación, números, eliminar etiquetas HTML, símbolos, cambio a minúsculas.
* **Tokenización**: proceso de dividir la cadena de texto en unidades más pequeñas llamadas **tokens** usando el espacio como seperardor**.** Los tokens incluyen palabras, caracteres y subpalabras.



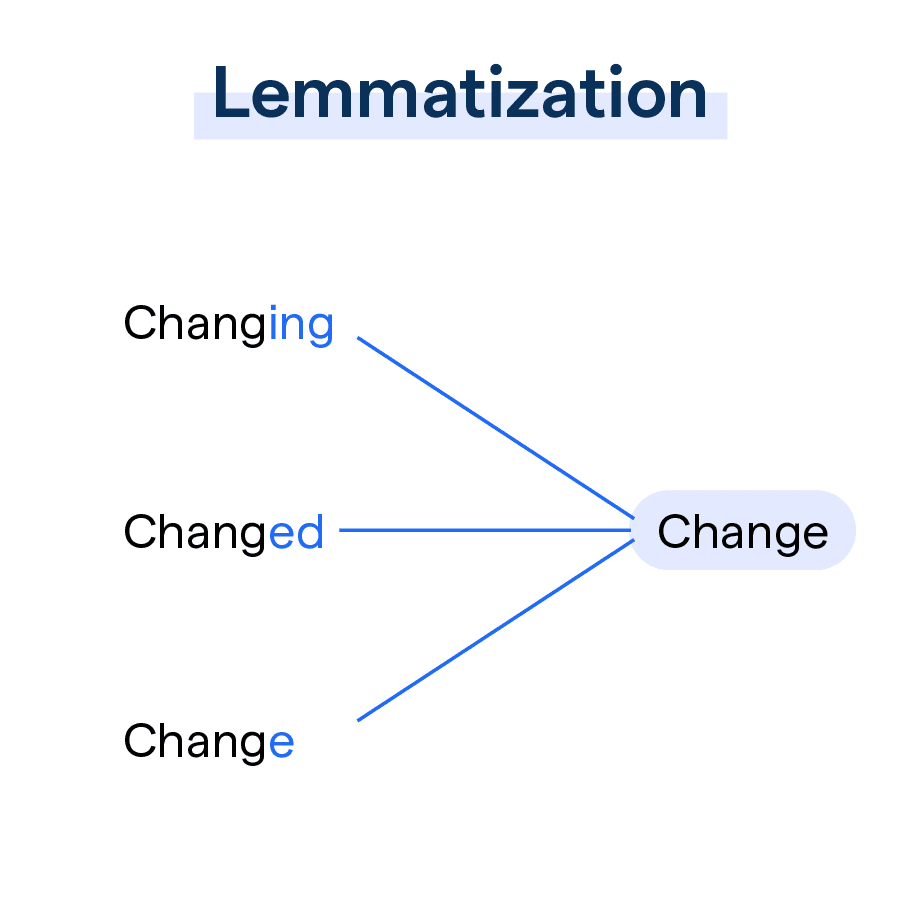
**Fig 3 - Ejemplo de tokenización de una frase**

* **Stemming**: una misma palabra se puede encontrar de diferentes formas dependiendo del género, del número, la persona etc. El stemming designa el simple proceso heurístico de cortar el final de las palabras para mantener solo la raíz de la palabra.



**Fig 4 - Ejemplo de stemming**

* **Lematización**: consiste en realizar la misma tarea que el stemming pero utilizando un vocabulario y un análisis minucioso de la construcción de las palabras. La lematización permite eliminar únicamente las terminaciones inflexibles y de ese modo aislar la forma canónica de la palabra, es decir, el lema.



**Fig 5 - Ejemplo de lematización**

Para poder aplicar métodos de Machine Learning a los problemas relativos al lenguaje natural, es indispensable transformar los datos textuales en datos digitales.

Alguna de las técnicas clásicas de PLN utiliza el **tf-idf** (Term Frequency-Inverse Document Frequency) como entrada para modelos clásicos de *Machine Learning* como pueden ser las máquinas de vectores soporte y los árboles de decisión.

Por un lado, el *Machine Learning* o Aprendizaje Automático es un subcampo de la Inteligencia Artificial el cual tiene como objetivo otorgar a las máquinas la habilidad de aprender sin ser explícitamente programadas (Arthur Samuel, 1959) [4]. O desde un punto más moderno (Tom Mitchell, 1997) [4]: “Un programa de computador se dice que aprende de la experiencia *E* con respecto a alguna clase de tareas de *T*, medido por *P*, mejora con la experiencia *E*.”

Dentro del *Machine Learning* existen diferentes tipos de aprendizaje, entre ellos los más comunes son el aprendizaje supervisado, en el cual se aprende una función a través de etiquetas y de los datos, y el no supervisado el cual se aprende una función a través de los datos.

Por otro lado, el TF-IDF es una medida numérica que expresa cuán relevante es una palabra para un documento en concreto dentro de un conjunto de documentos. Este método consiste en contar el número de ocurrencias de los tokens en el corpus para cada texto, que luego se divide por el número total de ocurrencias de esos mismos tokens en todo el corpus. Su formulación matemática es la siguiente:

Siendo ***tf (t, d)*** la frecuencia del término *t* en un documento *d:*

Donde:

* es el número de veces que aparece *t* en un documento *d*.
* es el número total de términos en el documento *d*

Siendola frecuencia de documento inversa, es decir, una medida de cuanta información proporciona una palabra, sobre como de común o específica es en todos los documentos.

Donde:

* *N* es el número total de documentos en el corpus,
* es el número de documento en los que el término *t* aparece.

Por tanto el valor de *tf-idf*  de las palabras de un texto, sirve como entrada para los modelos de *Machine Learning* tales como las Máquinas de Soporte Vectorial, Árboles de Decisión o Regresión Logística entre otros. Si bien estos modelos logran un buen resultado a la hora de resolver los problemas de PLN y más en concreto de clasificación de texto, el estado del arte viene marcado por distintos modelos basados en *Deep Learning*, entre los que destacan:

* **CNN,** Redes Neuronales Convolucionales
* **RNN,** Redes Neuronales Recurrentes.
* **Modelos con mecanismos de atención.**
* **Redes de memoria aumentada.**
* **Transformers y Modelos de lenguaje pre-entrenados.**

## **2.2.1 Transformers**