



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escola Tècnica  
Superior d'Enginyeria  
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica  
Universitat Politècnica de València

# **Adaptación de modelos de lenguaje grandes para la generación de lenguaje natural a partir de palabras clave en sistemas aumentativos y alternativos de comunicación**

**TRABAJO FIN DE GRADO**

Grado en Ciencia de Datos

*Autor:* Silvia Alegre Villa

*Tutor:* Jorge Civera Saiz

Curso 2023-2024



# Resum

Els Sistemes Augmentatius i Alternatius de Comunicació (SAAC) son eines essencials per a facilitar la comunicació de les persones amb dificultats en la utilització del llenguatge. Aquest sistema permeten a l'usuari la selecció de pictogrames associats a paraules clau que conformaran l'oració que es desitja comunicar. Posteriorment, aquesta oració pot ser sintetitzada amb veu humana. Els recents avanços en l'àrea del processament del llenguatge natural i, en concret, la proliferació de models de llenguatge grans ofereixen noves perspectives per a millorar els SAAC. En particular, aquest treball explorarà com aquests models de llenguatge poden millorar l'expressivitat de la comunicació dels SAAC quan s'utilitzen per a la generació de llenguatge natural a partir de les paraules clau (pictogrames) seleccionades per l'usuari. D'aquesta manera, aquest treball evaluarà el rendiment d'aquests models quan són adaptats per a la seua integració en els SAAC. Aquesta evaluació es durà a terme utilitzant conjunts de test reals en espanyol i anglès extrets del portal del Centre Aragonés per a la Comunicació Augmentativa i Alternativa.

**Paraules clau:** ????, ?????????, ????, ?????????????????

---

# Resumen

Los Sistemas Aumentativos y Alternativos de Comunicación (SAAC) son herramientas vitales para facilitar la comunicación de las personas con dificultades en la utilización del lenguaje. Estos sistemas permiten al usuario la selección de pictogramas asociados a palabras clave que conformarán la oración que se desea comunicar. Posteriormente, esta oración puede ser sintetizada con voz humana. Los recientes avances en el área del procesamiento de lenguaje natural y, en concreto, la proliferación de modelos de lenguaje grandes ofrece nuevas perspectivas para mejorar los SAAC. En particular, este trabajo explorará cómo estos modelos de lenguaje pueden mejorar la expresividad de la comunicación de los SAAC cuando se utilizan para la generación de lenguaje natural a partir de las palabras clave (pictogramas) seleccionadas por el usuario. De esta forma, este trabajo evaluará el rendimiento de estos modelos cuando son adaptados para su integración en los SAAC. Esta evaluación se llevará a cabo utilizando conjuntos de test reales en español e inglés extraídos del portal del Centro Aragonés para la Comunicación Aumentativa y Alternativa.

**Palabras clave:** ?????, ???, ?????????????????

---

# Abstract

Augmentative and Alternative Communication (AAC) systems are vital tools for facilitating communication for individuals with difficulties using language. These systems allow users to select pictograms associated with key words that will form the sentence that is wished to communicate. Then, the sentence can be synthesized with a human voice. Recent advances in the field of natural language processing, and specifically the proliferation of large language models, offer new perspectives for improving AAC systems. In particular, this work will explore how these language models can enhance the expressiveness of AAC communication when used to generate natural language from the key words (pictograms) selected by the user. In this way, this work will evaluate the performance of these models when adapted for integration into AAC systems. This evaluation will be carried out using real test sets in spanish and english extracted from the portal of the Aragonese Center for Augmentative and Alternative Communication.

**Key words:** ?????, ????? ?????, ?????????????

---



# Índice general

Índice general	V
Índice de figuras	VII
Índice de tablas	VII
<hr/>	
<b>1 Introducción</b>	<b>1</b>
1.1 Motivación . . . . .	1
1.2 Objetivos . . . . .	2
1.3 Estructura de la memoria . . . . .	2
<b>2 Fundamentos</b>	<b>3</b>
2.1 Aprendizaje automático . . . . .	3
2.2 Redes neuronales . . . . .	4
2.2.1 Redes neuronales para secuencias de texto . . . . .	5
2.2.2 Atención . . . . .	5
2.3 Transformers . . . . .	6
2.4 Procesamiento del lenguaje natural . . . . .	6
2.5 Modelos de lenguaje grandes (GPT, Gemma, Llama-3) . . . . .	6
2.5.1 Gemma . . . . .	6
2.5.2 Llama-3 . . . . .	6
<b>3 Adaptación de modelos de lenguaje grandes (PEFT: LoRA)</b>	<b>7</b>
<b>4 Resultados experimentales</b>	<b>9</b>
4.1 Conjunto de datos . . . . .	9
4.2 Medidas de evaluación . . . . .	9
4.3 Resultados experimentales . . . . .	9
<b>5 Conclusiones</b>	<b>11</b>
<b>Bibliografía</b>	<b>13</b>
<hr/>	
Apéndices	
<b>A Configuració del sistema</b>	<b>15</b>
A.1 Fase d'inicialització . . . . .	15
A.2 Identificació de dispositius . . . . .	15
<b>B ??? ???????????? ????</b>	<b>17</b>



## Índice de figuras

---

1.1	Comunicador AsTeRISCS Grid desarrollado por ARASAAC . . . . .	2
2.1	Ejemplo de red neuronal . . . . .	4

## Índice de tablas

---





---

---

# CAPÍTULO 1

## Introducción

---

ESCRIBIR BIEN !!! En este primer capítulo introductorio presentamos la motivación y objetivos del trabajo. También explicaremos cómo será la estructura del contenido de este.

### 1.1 Motivación

---

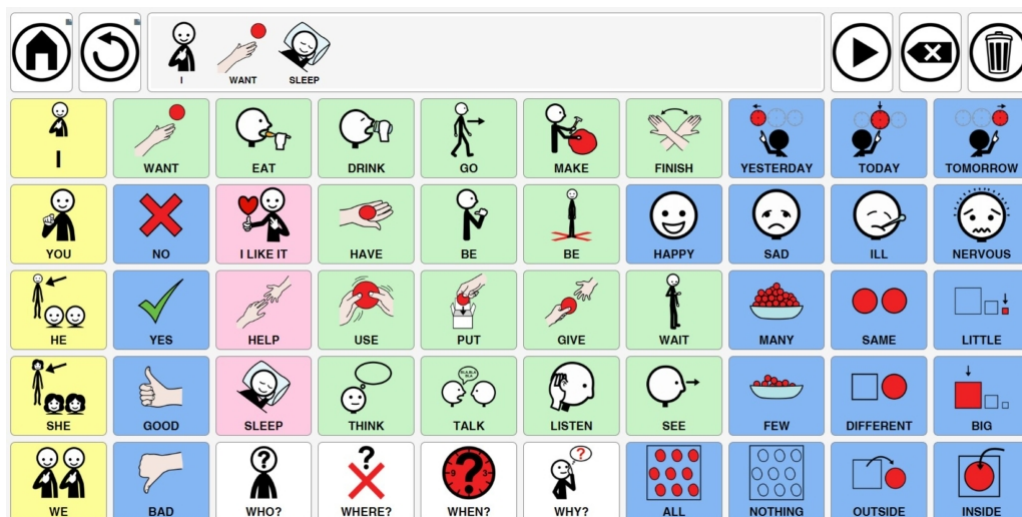
La comunicación y el lenguaje son dos pilares fundamentales de la sociedad actual, pues constituyen la base de las relaciones interpersonales, permitiendo el intercambio de ideas e información. Gracias a ello podemos transmitir a los demás nuestros pensamientos, emociones y necesidades, permitiéndonos participar en la vida en sociedad. Sin embargo, para algunas personas, el hecho de comunicarse de manera satisfactoria puede resultar realmente complicado debido a distintas causas. Es aquí donde entran en juego los Sistemas Aumentativos y Alternativos de Comunicación (SAAC).

Tal y como explica el Centro Aragonés para la Comunicación Aumentativa y Alternativa (ARASAAC), “los Sistemas Aumentativos y Alternativos de Comunicación son formas de expresión distintas al lenguaje hablado que tienen como objetivo aumentar el nivel de expresión y/o compensar las dificultades de comunicación y lenguaje de las personas que tienen dificultades en este ámbito”. Hay distintas razones por las cuales una persona podría necesitar utilizar un SAAC. Entre ellas encontramos la parálisis cerebral, la discapacidad intelectual, los trastornos del espectro autista, algunas enfermedades neurológicas, las distrofias musculares o las afasias, entre otras.

Aunque hay muchos tipos de SAAC, todos se caracterizan por estar basados en sistemas de símbolos, ya sean gráficos (fotografías, dibujos, pictogramas, palabras o letras) o gestuales (mímica o símbolos manuales). En este trabajo nos centraremos en los comunicadores electrónicos. Los comunicadores electrónicos son herramientas tecnológicas que pueden ser utilizados en cualquier tipo de dispositivo electrónico. Por lo general, consisten en un tablero donde aparecen distintos símbolos gráficos (pictogramas) que representan palabras.

Estos pictogramas pueden ser organizados y adaptados dependiendo de las necesidades de cada persona, permitiendo que cada usuario añada aquellos que necesite. El funcionamiento es simple: el usuario clicca sobre los pictogramas y el programa se encarga de dar la salida del mensaje, generalmente en forma de habla digitalizada o en formato escrito.

Así, estas herramientas resultan verdaderamente útiles para solventar los problemas de comunicación de las personas. Sin embargo, presentan una limitación que puede afectar a la fluidez y al nivel de expresividad con que se realiza la comunicación: la dificultad



**Figura 1.1:** Comunicador AsTeRISCS Grid desarrollado por ARASAAC

para conjugar las frases de manera adecuada, pues para facilitar y simplificar el uso de la herramienta, en este tipo de tableros se suelen incluir solamente palabras clave en su forma simple, sin distinción de número, género o tiempo verbal.

Algunos de los comunicadores que existen actualmente en el mercado ya emplean diferentes métodos para abordar este problema, pero existe todavía un amplio margen de mejora. Los modelos de lenguaje grandes prometen ofrecer buenos resultados en este área.

## 1.2 Objetivos

Los objetivos de este proyecto son los siguientes:

1. Investigar sobre los enfoques actuales para la generación de frases en el sector de los SAAC.
2. Implementar y evaluar modelos de lenguaje grandes para este tipo de herramientas.
3. Comparar los resultados que se obtienen con los modelos de lenguaje grandes respecto a otros comunicadores que permiten la generación de frases que se encuentran en el mercado, utilizando para ello las métricas de BLEU y COMET.

## 1.3 Estructura de la memoria

????? ?????????????? ?????????????? ?????????????? ?????????????? ??????????????

---

## CAPÍTULO 2

# Fundamentos

---

REESCRIBIR: Los modelos de lenguaje se encuadran dentro de las técnicas de procesamiento de lenguaje natural, las cuales forman parte del ámbito del aprendizaje automático. Antes de profundizar en las tareas específicas que realizan los modelos de lenguaje grandes, debemos introducir los conceptos básicos del aprendizaje automático, del aprendizaje profundo y del procesamiento de lenguaje natural. Así, en este capítulo abordaremos las distintas tareas y enfoques del aprendizaje automático, proporcionando el contexto necesario para comprender el funcionamiento y las aplicaciones de los modelos de lenguaje.

### 2.1 Aprendizaje automático

---

El aprendizaje automático (en inglés, *machine learning*) es una disciplina dentro de la inteligencia artificial que se centra en el desarrollo y estudio de algoritmos y modelos que permiten que los sistemas puedan realizar tareas específicas sin haber sido explícitamente programados para ello. Este término fue acuñado por Arthur Samuel en el año 1959, quien creó uno de los primeros programas exitosos de esta disciplina, conocido como *the Samuel Checkers-playing Program* [?] (el programa de juego de damas de Samuel).

Tom Mitchell [1] define el proceso de aprendizaje de los programas en el campo del *machine learning* de la siguiente manera:

*"Se dice que un programa aprende de la experiencia  $E$  con respecto a alguna clase de tarea  $T$ , y medida de rendimiento  $P$ , si su rendimiento en tareas en  $T$ , medido por  $P$ , mejora con la experiencia  $E$ ."*

Aunque la idea principal del aprendizaje automático es esta, encontramos diferentes enfoques para el entrenamiento de los algoritmos, dependiendo del tipo de tarea que se quiera llevar a cabo. Estos enfoques son: el aprendizaje supervisado, cuyo objetivo es, a partir de unos datos de entrenamiento, encontrar la función  $f$  que realice el mejor mapeo posible entre un conjunto de entradas  $X$  y sus salidas correspondientes  $Y$ , de manera que  $(X, Y) = (X, f(Y))$ ; el aprendizaje no supervisado, que trata de modelar la estructura subyacente de un conjunto de datos para identificar relaciones y patrones, permitiendo así un entendimiento más profundo de los mismos; el aprendizaje semi-supervisado, que cae entre el supervisado y el no supervisado y utiliza una porción de datos etiquetados y no etiquetados y, por último, el aprendizaje por refuerzo, donde el algoritmo aprende a través de retroalimentaciones que va recibiendo, ajustando sus acciones con el objetivo de maximizar una recompensa acumulada a lo largo del tiempo. La tarea principal de este trabajo se realizará utilizando las técnicas del aprendizaje supervisado.

Otro concepto importante dentro del aprendizaje automático es el aprendizaje profundo (*deep learning*). El aprendizaje profundo es subconjunto dentro del aprendizaje automático que emplea algoritmos basados en redes neuronales. Dentro de este encontramos métodos como las redes neuronales profundas, las redes neuronales recurrentes, las redes neuronales convolucionales y los transformers, entre otros. Estos métodos tienen aplicaciones significativas en una gran variedad de ámbitos, entre los que se encuentra el procesamiento de lenguaje natural, disciplina en la que se enmarca este trabajo. En las siguientes secciones explicaremos con detalle los conceptos de redes neuronales y transformers.

## 2.2 Redes neuronales

El primer modelo de red neuronal artificial fue creado en 1943 por Warren McCulloch y Walter Pitts. Este era un modelo simplificado que imitaba el comportamiento del cerebro humano, y dio inicio a un proceso de investigación de las redes neuronales desde dos enfoques distintos: uno centrado en los procesos biológicos del cerebro y otro en la aplicación de las redes neuronales para la inteligencia artificial. Las redes neuronales modernas, aunque inspiradas en estas primeras ideas, han evolucionado significativamente y ya no se basan directamente en las inspiraciones biológicas iniciales.

Las redes neuronales modernas están compuestas por pequeñas unidades de cómputo conocidas como neuronas o nodos, conectadas entre sí a través de enlaces para permitir la transmisión de señales entre estas. Generalmente el flujo de datos se inicia en los nodos de la capa de entrada (*input layer*), pasa por una o más capas intermedias conocidas como capas ocultas (*hidden layers*), y llega a los nodos de la capa de salida (*output layer*).

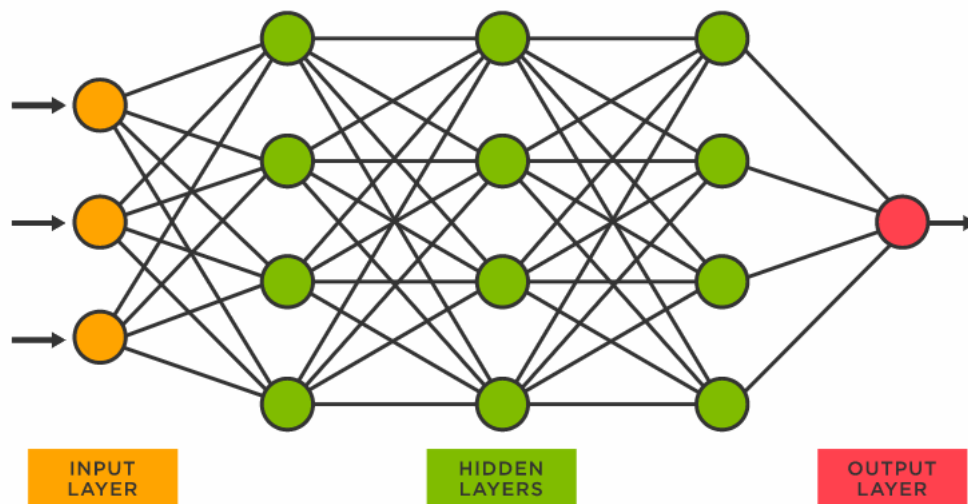


Figura 2.1: Ejemplo de red neuronal

En la primera de estas capas los datos entran al modelo, en forma de vector. A cada una de las neuronas de la capa de entrada le corresponde un elemento del vector, sobre el cual realizará algún tipo de cálculo para luego enviar información a la siguiente neurona. Cada uno de los enlaces tiene asociado un peso, el cual ponderará el resultado anterior. Así, el valor que llega a la siguiente neurona es la suma ponderada de las entradas anteriores por los pesos de los enlaces, al cual se le puede añadir también un umbral, también conocido como *bias term*, para poder controlar el nivel de activación de las neuronas. Da-

do un conjunto de entradas  $x_1, \dots, x_n$ , los correspondientes pesos de los enlaces  $w_1, \dots, w_n$  y un término de *bias*  $b$ , la suma ponderada  $z$  se calcula:

$$z = b + \sum_i w_i x_i \quad (2.1)$$

En lugar de utilizar como output  $z$ , que es una función lineal de  $x$ , se aplica una función no lineal  $f$  a  $z$ , conocida como función de activación, y es el resultado de esta el que llegará a la siguiente neurona. Entre las funciones de activación más comunes se encuentran la función sigmoide, la tangente hiperbólica (*tanh*) y la rectificadora lineal (*ReLU*).

### 2.2.1. Redes neuronales para secuencias de texto

En esta sección presentaremos dos tipos de redes neuronales que se utilizan para tareas con secuencias de texto.

#### Redes neuronales recurrentes (RNN)

Las redes neuronales recurrentes (en inglés, *recurrent neural networks* (RNN)) son un tipo de red neuronal que mapean desde un conjunto de entrada a otro de salida de manera que esta predicción  $y_t$  depende no solo de la de entrada  $x_t$  sino también del estado oculto del sistema,  $h_t$ , que se actualiza con el tiempo a medida que se procesa los datos. Este tipo de modelos se pueden utilizar para la generación, clasificación y traducción de texto.

#### Redes neuronales convolucionales (CNN)

El funcionamiento de las redes neuronales convolucionales (*convolutional neural networks* (CNN)) se basa en el cálculo de una función de un vecindario local para cada una de las entradas utilizando pesos compartidos. Son una buena alternativa a las redes neuronales recurrentes, pues son sustancialmente más sencillas de entrenar, ya que no necesitan mantener el estado oculto a largo plazo. Se pueden utilizar en tareas de clasificación y de generación de texto.

### 2.2.2. Atención

En las redes neuronales clásicas, el cálculo de cada capa mediante una combinación lineal de los vectores de entrada y los correspondientes pesos, seguida de la aplicación de una función no lineal, conocida como función de activación. El resultado de esta operación es  $Z = f(XW)$ , donde  $X$  es el vector de *inputs*,  $W$  es el vector de pesos,  $f$  es la función de activación y  $Z$  son los *outputs* producidos.

Sin embargo, podemos hacer que los modelos sean aún más flexibles si permitimos que los pesos dependan de los *inputs*. Esto se representa como  $Z = f(XW(X))$ . A este tipo de interacción multiplicativa, donde los pesos varían en función de los *inputs*, es lo que se conoce como atención.

De manera más general, el mecanismo de atención puede describirse mediante la fórmula  $Z = f(VW(Q, K))$ . En este contexto,  $Q$  (*queries*) es un conjunto derivado de  $X$  que se utiliza para describir lo que se busca con cada *input*,  $K$  (*keys*) es otro conjunto derivado de  $X$  que se usa para describir qué contiene cada vector de entrada y  $V$  (*values*)

es un conjunto también derivado de  $X$  que se utiliza para describir cómo cada input debe ser transmitido hasta el output.

## 2.3 Transformers

---

En esta sección presentaremos la arquitectura de transformers.

Los modelos basados en transformers son modelos del tipo *seq2seq* (tanto la entrada como la salida son secuencias de texto) que utilizan el concepto de atención definido en la sección anterior. La arquitectura de transformers fue introducida en el artículo "*Attention is All You Need*" en 2017, y supuso una gran revolución en el campo del procesamiento de lenguaje natural y otras áreas del aprendizaje automático.

Los transformers destacan por su capacidad para manejar dependencias a largo plazo en las secuencias de datos de manera eficiente, superando a las redes neuronales recurrentes y las redes neuronales convolucionales en muchas tareas. Además, utilizan mecanismos de autoatención (*self-attention*), que permiten ponderar la importancia de las diferentes partes de la secuencia de entrada al generar cada parte de la secuencia de salida.

## 2.4 Procesamiento del lenguaje natural

---

## 2.5 Modelos de lenguaje grandes (GPT, Gemma, Llama-3)

---

Los modelos de lenguaje grandes (en inglés, *Large Language Models* o LLM) son modelos de aprendizaje profundo basados en redes neuronales con miles de millones de parámetros, que utilizan la arquitectura de transformers. Estos modelos pueden ser entrenados con grandes cantidades de texto, permitiéndoles realizar una gran cantidad de tareas en el ámbito del procesamiento de lenguaje natural.

Inicialmente, los LLM operaban de manera secuencial, basando sus predicciones en las distribuciones de probabilidad de las palabras dentro de un texto. Sin embargo, este enfoque tenía la limitación de no considerar el contexto más amplio en el que aparecen las palabras, así como sus distintos significados y asociaciones. Los avances en la arquitectura de las redes neuronales y, especialmente, la aparición de los transformers, han representado una gran evolución hacia modelos de lenguaje grandes mucho más avanzados, capaces de procesar simultáneamente enormes cantidades de texto y permitiendo establecer relaciones más sólidas entre las palabras y el contexto en que aparecen.

### 2.5.1. Gemma

### 2.5.2. Llama-3

---

## CAPÍTULO 3

# Adaptación de modelos de lenguaje grandes (PEFT: LoRA)

---

???? ????????????? ????????????? ????????????? ????????????? ?????????????





---

---

## CAPÍTULO 4

# Resultados experimentales

---

### 4.1 Conjunto de datos

---

### 4.2 Medidas de evaluación

---

### 4.3 Resultados experimentales

---

???? ????????????? ????????????? ????????????? ????????????? ?????????????



---

---

## CAPÍTULO 5

# Conclusions

---

????? ?????????????? ?????????????? ?????????????? ?????????????? ??????????????



## Bibliografía

---

- [1] Tom M Mitchell and Machine Learning. McGraw-hill. *New York*, pages 154–200, 1997.



---

---

## APÉNDICE A

# Configuració del sistema

---

???? ????????????? ????????????? ????????????? ????????????? ?????????????

### A.1 Fase d'inicialització

---

???? ????????????? ????????????? ????????????? ????????????? ?????????????

### A.2 Identificació de dispositius

---

???? ????????????? ????????????? ????????????? ????????????? ?????????????





---

---

## APÉNDICE B

??? ?????????????????? ?????

---

???? ????????????????? ????????????????? ????????????????? ????????????????? ?????????????????