



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

**Escuela Técnica Superior de Ingeniería Informática**

**Proyecto de fin de grado**

**Optimización y evaluación de sistemas de traducción  
automática para la asistencia a la gestión de recursos  
humanos en un contexto multilingüe**

**Grado en ingeniería informática**

**Curso académico: 2024-2025**

Juan Castelló Beltrán

Tutor: Jorge Civera Saiz

10 de junio de 2025

# Índice

<b>1</b>	<b>Introducción y motivación</b>	<b>3</b>
<b>2</b>	<b>Estado del arte</b>	<b>5</b>
<b>3</b>	<b>Descripción del conjunto de datos</b>	<b>7</b>
<b>4</b>	<b>Metodología</b>	<b>9</b>
<b>5</b>	<b>Resultados</b>	<b>13</b>
<b>6</b>	<b>Conclusiones</b>	<b>16</b>

# 1. Introducción y motivación

## ■ Contexto actual

- Globalización y empresas multinacionales.
- Gestión de RRHH.
- Dificultad en la identificación de talento:
  - Diversidad de idiomas.
  - Gran cantidad de datos.



# Introducción y motivación

## ■ Solución propuesta por la empresa (NTT Data / UNV)

- Utilización de inteligencia artificial (IA) para el procesamiento de datos.
- Implementación de sistemas de búsqueda inteligente de candidatos.
- Estandarización de información a un idioma común (inglés).

## ■ Motivación y Objetivo del TFG

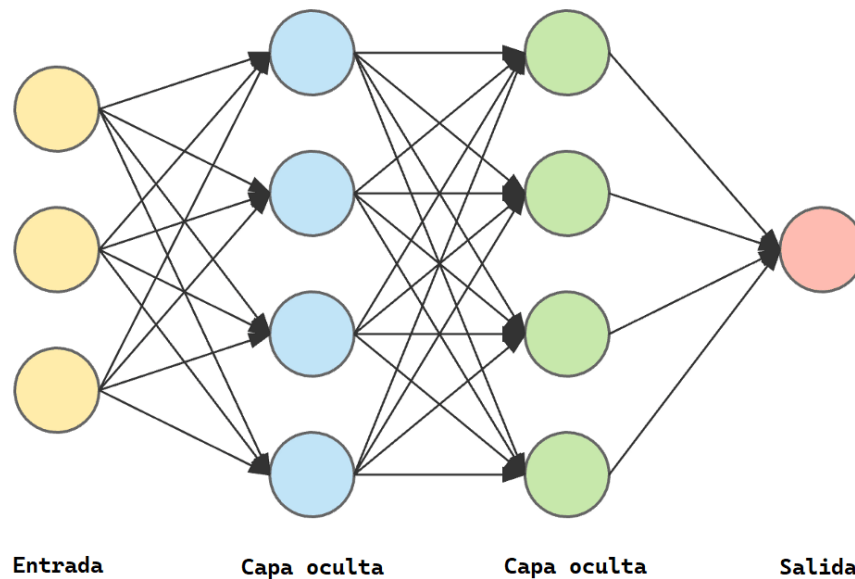
- Optimizar y evaluar el módulo de traducción automática (TA).
- Uso de modelos de lenguaje grandes (LLMs).
- Buscar un equilibrio entre calidad, y coste temporal y económico.

## 2. Estado del arte

### ■ Aprendizaje automático (ML)

- Aprendizaje supervisado.

### ■ Redes neuronales



- Arquitectura Transformer: mecanismo de autoatención

## 2 Estado del arte

### ■ Modelos de Lenguaje Grandes (LLMs)

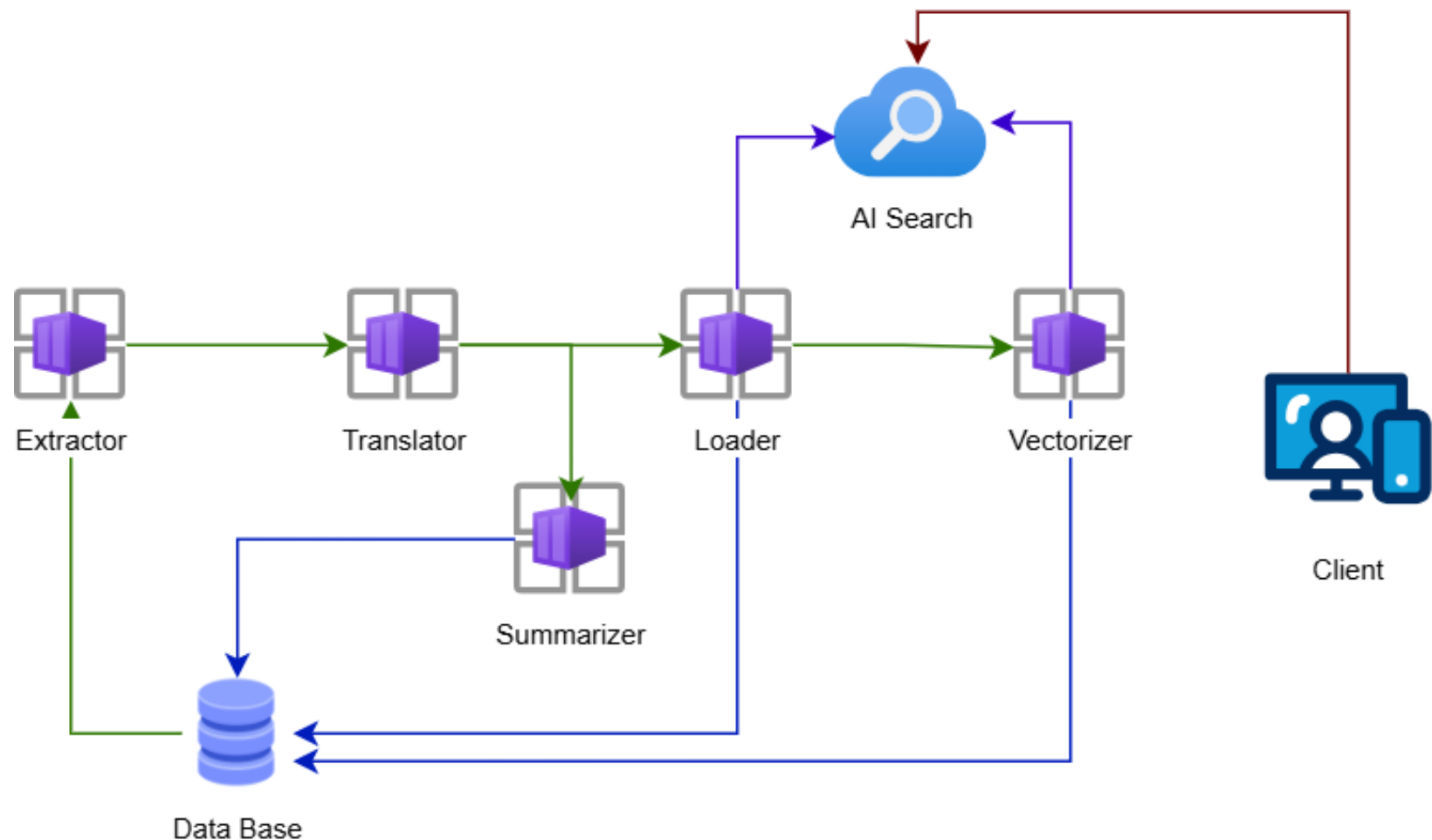
- Basados en la arquitectura Transformer.
- Preentrenamiento con gran cantidad de datos.
- Adaptación a tareas concretas: **zero-shot**, **few-shot** y **fine-tuning**.
- Capacidad de generación y comprensión de texto.

### ■ Prompt engineering

- Clave para guiar el comportamiento de los LLMs.
- Impacta directamente en la calidad de la traducción.

### 3. Descripción del conjunto de datos

- **Tipos de datos:** Perfiles - Documentos de Asignación (DoA) - Asignaciones
- **Aplicación completa**



# 3 Descripción del conjunto de datos

## ■ Selección y preproceso

- Filtrado de datos.
- Extracción de frases para la traducción.
- Selección de 1.000 frases (66 % perfiles, 33 % asignaciones, 1 % DoAs)

## ■ Generación de traducciones supervisadas

- Traducción inicial con DeepL.
- Revisión y posesición manual exhaustiva.
- Terminología específica.



## 4. Metodología

### ■ Evaluación de la traducción automática

- Evaluación manual (subjetiva).
- Evaluación automática (objetiva).

### ■ Tipos de métricas automáticas

- Basadas en coincidencias léxicas (BLEU).
- Basadas en similitud semántica (COMET).

# Metodología: evaluación automática

BLEU	COMET
Comparación de n-gramas.	Comparación de word embeddings.
<b>Ventajas:</b> <ul style="list-style-type: none"><li>■ Rápido y económico.</li><li>■ Funcionamiento simple.</li><li>■ Independiente del idioma.</li></ul>	<b>Ventajas:</b> <ul style="list-style-type: none"><li>■ Alta sensibilidad semántica.</li><li>■ Fuerte correlación con juicio humano.</li><li>■ Considera la fidelidad al texto fuente.</li></ul>
<b>Desventajas:</b> <ul style="list-style-type: none"><li>■ Problemas con frases largas.</li><li>■ Problemas con sinónimos.</li></ul>	<b>Desventajas:</b> <ul style="list-style-type: none"><li>■ Mayor complejidad computacional.</li><li>■ Dependencia del modelo base.</li></ul>
<ul style="list-style-type: none"><li>■ 30 - 40: Buenas traducciones</li><li>■ 40 - 50: Alta calidad</li><li>■ 50 - 60: Calidad muy alta</li></ul>	<ul style="list-style-type: none"><li>■ <math>&lt; 85</math>: Alta calidad</li><li>■ <math>\geq 85</math>: Calidad muy alta</li></ul>

# Metodología: LLMs

- Criterio de selección: Diversidad en rendimiento, coste y accesibilidad.
- Modelos de Azure OpenAI:
  - Familia ChatGPT: GPT 4, GPT 4 mini.
- Modelos locales / Open-Source:
  - Familia Gemma3: Gemma3 4b, Gemma3 4b translator, Gemma3 12b
  - Familia DeepSeek: DeepSeek 7b, DeepSeek 8b.

# Metodología: prompt engineering

- Se han utilizado 3 prompts distintos.
- **Prompt 1:** descripción de tarea a realizar (traducción)  
*Traduce estas frases del español al inglés.*
- **Prompt 2:** introduce un contexto según el tipo de documento a traducir  
*Actúa como un traductor que analiza las frases con la información de perfiles y tradúcelos del español al inglés.*
- **Prompt 3:** añade un ejemplo de traducción (one-shot)  
*Actúa como un traductor que analiza las frases con la información de perfiles y tradúcelos del español al inglés como el ejemplo siguiente:  
Agradezco la oportunidad de considerar mi solicitud.  
I appreciate the opportunity to consider my application.*

## 5. Resultados

### ■ Rendimiento Global (BLEU y COMET)

Modelo LLM	Prompt 1		Prompt 2		Prompt 3	
	BLEU	COMET	BLEU	COMET	BLEU	COMET
GPT 4	56.8	87.8	<b>58.8</b>	<b>88.0</b>	57.9	<b>88.0</b>
GPT 4 mini	<b>58.3</b>	<b>88.0</b>	57.6	87.9	<b>58.3</b>	87.9
Gemma3 4b	47.6	86.3	48.4	86.5	51.9	86.0
Gemma3 4b translator	<b>58.0</b>	<b>86.5</b>	<b>57.0</b>	<b>86.6</b>	<b>57.7</b>	<b>86.1</b>
Gemma3 12b	48.0	86.9	47.0	86.8	52.1	87.1
DeepSeek 7b	29.8	81.4	30.7	81.8	28.1	80.9
DeepSeek 8b	37.8	83.4	37.9	83.5	37.7	83.6

### ■ Impacto del Prompt Engineering

- La familia GPT muestra los mejores resultados.
- Gemma3 4b translator ofrece los resultados más competitivos a GPT
- Balance entre complejidad del prompt y potencia del LLM.

# Resultados: Análisis de coste temporal y económico

Modelo	Coste temporal (horas)	Coste económico (€)
GPT 4	0.2	0.52
GPT 4 mini	0.2	0.03
Gemma 3 4b	1.1	0.03
Gemma 3 4b-translation	1.1	0.03
Gemma 3 12b	2.1	0.06
DeepSeek 7b	17.7	0.53
DeepSeek 8b	20.0	0.61

- GPT 4 mini ofrece el mejor equilibrio entre calidad de traducción y costes
- Gemma3 4b translator y Gemma 3 12b son los modelos gratuitos más competitivos

# Resultados: Tipos de documentos

## ■ Rendimiento por tipo de documento

Modelo LLM	Asignaciones		Perfiles		DoAs	
	BLEU	COMET	BLEU	COMET	BLEU	COMET
<b>GPT 4 mini</b>	63.08	89.40	55.75	87.37	58.67	90.84
<b>Gemma3 12b</b>	60.10	88.78	<b>47.75</b>	<b>86.32</b>	51.33	88.75
<b>DeepSeek 8b</b>	44.92	85.82	<b>33.70</b>	<b>82.58</b>	38.39	85.76

- Variaciones significativas de calidad según el tipo de documento.
- GPT 4 mini es más robusto con calidad alta en diferentes documentos.

## 6. Conclusiones

### ■ Rendimiento de los LLMs en TA para RRHH

- LLMs actuales (GPT) altamente efectivos para TA en el dominio de RRHH.
- Mejora de eficiencia y precisión en gestión de documentación multilingüe.

### ■ Resultados

- Búsqueda de equilibrio entre calidad y eficiencia temporal/económica.
- Selección de modelos en función de prioridades empresariales.
- Repetir las ejecuciones locales con una máquina más potente

### ■ Relevancia del Prompt Engineering

- Diseño de prompt es crucial para optimizar el rendimiento de la TA.