



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

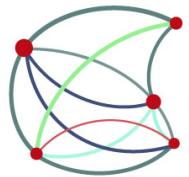
# Adaptación de modelos de lenguaje grandes para tareas de traducción automática

Sergio Madrid Pérez

Alfons Juan Císcar

Jorge Civera Saiz

Jorge Iranzo Sánchez



**MLLP**

Machine Learning  
and Language Processing

*DSIC*  
DEPARTAMENT DE SISTEMES  
INFORMÀTICS I COMPUTACIÓ

# Índice

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Fundamentos Teóricos</b>	<b>4</b>
<b>3</b>	<b>Modelos encoder-decoder</b>	<b>7</b>
<b>4</b>	<b>Adaptación de LLMs</b>	<b>10</b>
<b>5</b>	<b>Conclusiones</b>	<b>14</b>

# 1. Introducció

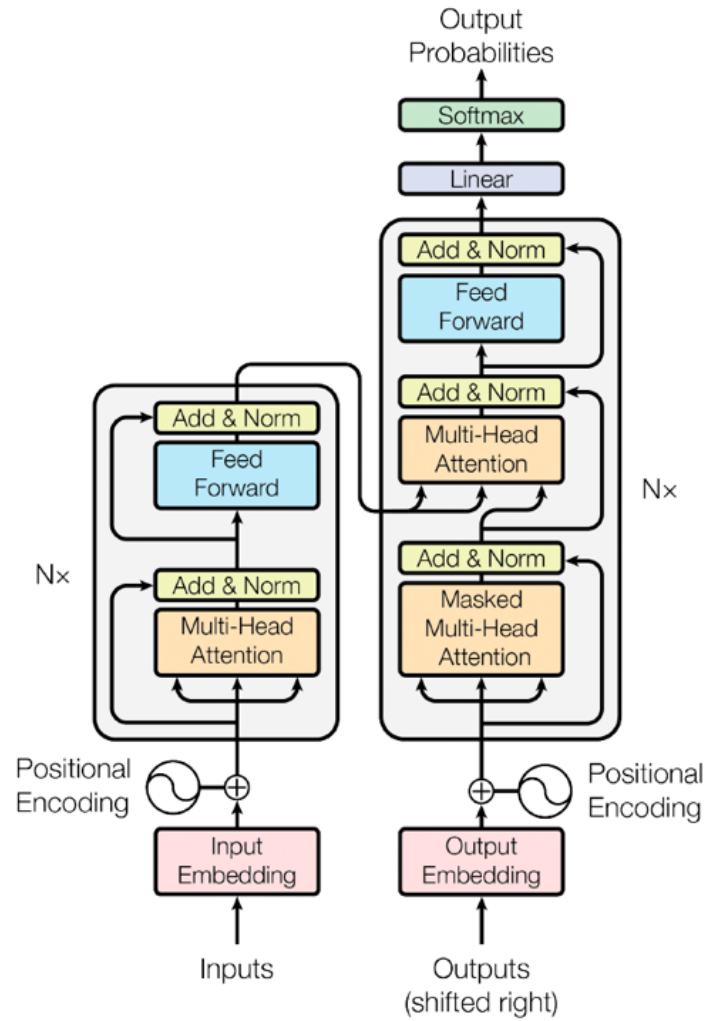
## Motivació

- Traducció automàtica (MT) se basa en models neuronals (Transformer)
- Models de llenguatge grans (LLMs) han revolucionat la IA
- LLMs molt exitosos en MT y, en general, en processament del llenguatge natural (NLP)

## Objectivos

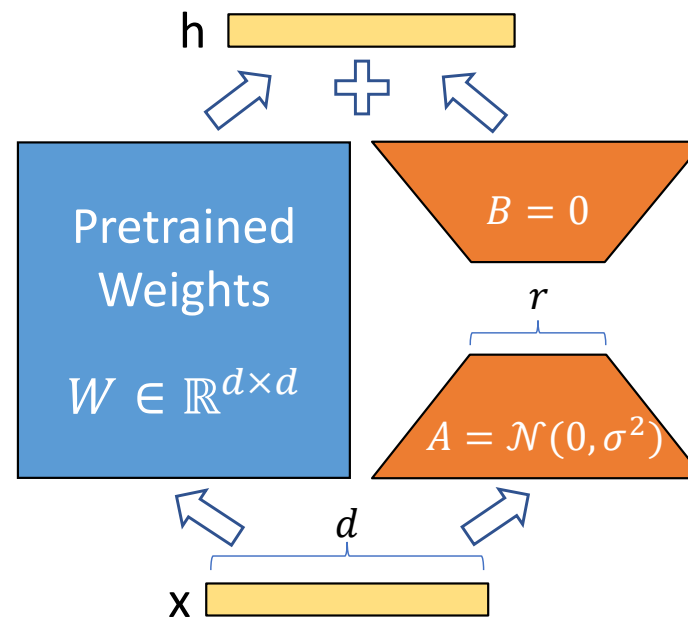
- Evaluar los modelos encoder-decoder más recientes diseñados para TA.
- Adaptar y evaluar LLMs decoder-only para TA.
- Evaluar las capacidades de aprendizaje en contexto de los LLMs para TA.

## 2. Fundamentos: Transformer



## 2. Fundamentos: LLMs

- Típicamente decoder de Transformer con gran número de parámetros
- Se entrenan con colecciones de datos muy extensas
- Adaptabilidad a diversas tareas NLP con gran éxito
- Uso de técnicas de ajuste eficiente de parámetros (PEFT) como LoRA



## 2. Fundamentos: neural MT

- Estado del arte: transformers (arquitectura encoder-decoder)
- Últimos resultados con LLMs comparables con el estado del arte
- Métricas de evaluación usuales: BLEU y COMET

### 3. Modelos encoder-decoder: Datasets

#### Evaluación

- INTERACT
- Europarl-ST

Nº de oraciones en cada conjunto de test

$en \rightarrow$	INTERACT	Europarl-ST
$es$	1405	1267
$de$	1399	1253

#### Entrenamiento

- Medline-WMT22
- Europarl-ST
- MuST-C

$en \rightarrow$	Sentences	Words	
		Source	Target
$es$	442.5 K	9.2 M	9.7 M
$de$	361.2 K	7.1 M	6.7 M

### 3. Modelos encoder-decoder: INTERACT

Model	LORA	Spanish		German	
		BLEU	COMET	BLEU	COMET
Google Translate	No	56.7	<b>87.6</b>	40.5	<b>86.5</b>
Helsinki-500M	No	55.6	85.4	37.2	80.9
Madlad-3B	No	55.8	85.7	<b>43.4</b>	83.5
NLLB-600M	No	55.3	86.1	37.3	82.2
NLLB-1.3B	No	55.9	86.2	39.3	82.9
NLLB-3.3B	No	56.3	86.3	41.1	83.5
NLLB-600M	Yes	56.0	86.4	38.2	83.0
NLLB-1.3B	Yes	57.2	87.1	41.2	84.5
NLLB-3.3B	Yes	<b>58.8</b>	87.5	43.1	85.2



### 3. Modelos encoder-decoder: Europarl-ST

Model	LORA	Spanish		German	
		BLEU	COMET	BLEU	COMET
Google Translate	No	48.1	<b>89.8</b>	34.4	<b>89.3</b>
Helsinki-500M	No	46.9	89.0	35.8	87.3
Madlad-3B	No	<b>49.0</b>	89.2	<b>38.9</b>	88.5
NLLB-600M	No	44.4	88.6	31.4	86.8
NLLB-1.3B	No	46.2	89.0	33.4	87.4
NLLB-3.3B	No	47.3	89.4	35.1	88.1
NLLB-600M	Yes	46.7	88.7	35.3	87.6
NLLB-1.3B	Yes	48.0	89.3	37.2	88.3
NLLB-3.3B	Yes	<b>49.0</b>	89.4	38.5	88.8

## 4. Adaptación de LLMs: INTERACT

Model	Spanish		German	
	BLEU	COMET	BLEU	COMET
Llama3-8B	<b>52.1</b>	<b>86.3</b>	<b>36.1</b>	<b>84.2</b>
Mistral-7B	50.6	86.2	34.7	83.8
Llama2-7B	51.0	86.2	33.5	83.4
Gemma-7B	50.8	85.8	34.7	83.9
Falcon-7B	49.5	86.0	33.4	83.1

## 4. Adaptación de LLMs: Europarl-ST

Model	Spanish		German	
	BLEU	COMET	BLEU	COMET
Llama3-8B	<b>47.5</b>	<b>89.5</b>	<b>35.6</b>	<b>88.5</b>
Mistral-7B	46.8	89.5	34.5	88.4
Llama2-7B	46.7	89.3	34.6	88.3
Gemma-7B	46.6	89.2	34.5	88.2
Falcon-7B	46.0	89.1	33.5	87.6

## 4. Comparación: NLLB-3.3B vs Llama3 en INTERACT

Model	LoRA	Spanish		German	
		BLEU	COMET	BLEU	COMET
NLLB-3.3B	Yes	<b>58.8</b>	<b>87.5</b>	<b>43.1</b>	<b>85.2</b>
Llama3-8B	Yes	52.1	86.3	36.1	84.2

## 4. Comparación: Ejemplos de traducciones NLLB-3.3 y Llama3

### Ejemplo 1

---

Source	Hopefully, we will in the future.
Reference	Con suerte, lo haremos en el futuro.
NLLB-3.3B	Con suerte, lo haremos en el futuro. (BLEU: 100)
Llama3-8B	Esperemos que en el futuro podamos hacerlo. (BLEU: 19.5)

---

### Ejemplo 2

---

Source	So, it has to be robust to the patient variations, and also to the <b>treatment delivery</b> .
Reference	Por lo tanto, tiene que ser robusto para las variaciones de los pacientes, y también para la <b>administración del tratamiento</b> .
NLLB-3.3B	Por lo tanto, tiene que ser robusto para las variaciones de los pacientes, y también para la <b>entrega del tratamiento</b> . (BLEU: 87.8)
Llama3-8B	Entonces, tiene que ser robusto a las variaciones del paciente, y también a la <b>administración del tratamiento</b> . (BLEU: 38.43)

---

# 5. Conclusiones

## Objetivos logrados

- Modelos MT encoder-decoder evaluados en INTERACT y Europarl-ST
- LLMs actuales adaptados y evaluados en INTERACT y Europarl-ST
- Comparativa: resultados con LLMs comparables al estado del arte

## Trabajo futuro

- Evaluar el rendimiento de estos modelos para pares de lenguas de bajos recursos.
- Emplear modelos más grandes y determinar si la mejora en las traducciones compensa el aumento en los recursos necesarios para desplegar estos modelos.
- Explorar diferentes estrategias de prompting que guíen al modelo hacia una mejor traducción.
- Estudiar la aplicación de los LLMs para sistemas de TA en streaming.