



## Becas colaboración curso 2023/2024

Fecha: 29 Mayo 2023

### Vicerrectorado de Investigación

Subcomisión de I+D+i

Propuesta del departamento *SISTEMAS INFORMÁTICOS Y COMPUTACIÓN*

**Núm Proyecto: 2023/32/00015**

#### Responsable

Casacuberta Nolla, Francisco

#### E-mail

fcn@prhlt.upv.es

#### Ext.

73515

#### Responsable

Paredes Palacios, Roberto

#### E-mail

rparedes@dsic.upv.es

#### Ext

73525

#### Título proyecto

Desarrollo de sistemas de traducción automática utilizando datos sintéticos generados a partir de grandes modelos de lenguaje pre-entrenados.

#### Valoración proyecto

4

#### Descripción proyecto

En este proyecto se plantea el estudio del uso de grandes modelos de lenguaje pre-entrenados (LLM), como los de la familia GPT, para generar nuevas traducciones sintéticas con las que entrenar el modelo de un sistema de traducción automática para tareas específicas. En primer lugar, a partir de un conjunto texto monolingüe se van a generar nuevos texto y traducciones con el LLM. Esta idea se puede llevar al extremo de no utilizar ningún conjunto de entrenamiento y solo una breve descripción del dominio en el que se pretende trabajar.

#### Actividades a realizar por el alumno

1. Estudiar el estado del arte de los grandes modelos de lenguaje para la generación de textos así como el de los sistemas de traducción automática.
2. Familiarización con el paquetes de software existentes como los de la familia GPT y openNMT.
3. Búsqueda de una tarea de traducción automática con sus correspondientes conjuntos de datos.
4. Generación de datos sintéticos utilizando los grandes modelos de lenguaje escogidos.
5. Construcción de un prototipo para la traducción automática basado en redes neuronales mediante el paquete openNMT.
5. Evaluación del prototipo construido.
6. Redacción de la memoria donde se describa el trabajo realizado así como las conclusiones que se puedan extraer y el trabajo futuro.

#### Localización de la actividad (Campus)



## Becas colaboración curso 2023/2024

*Fecha: 29 Mayo 2023*

Vera

### **Horario**

20 horas semanales ajustables a los horarios de clase