

# Una empresa del Parque Científico de la UMH crea modelos de lenguaje de IA

---

► La compañía Prompsit diseña un banco de información en varios idiomas que es masivo y libre

---

**R.A.**

■ Combinar grandes cantidades de datos, diferentes idiomas y computación de alto rendimiento para crear modelos de lenguaje y traducción potentes y eficientes, es el objetivo del proyecto High Performance Language Technologies (HPLT), en el que participa activamente la empresa Prompsit del Parque Científico de la Universidad Miguel Hernández de Elche (PCUMH). Esta iniciativa europea sobre tecnologías del lenguaje y la Inteligencia Artificial (IA) se enmarca en el programa Horizon 2030 y en ella participan, además de la compañía del PCUMH, cinco universidades y dos centros de supercomputación.

Concretamente, en el marco de este proyecto, Prompsit se ha encargado de diseñar y desarrollar un corpus multilingüe masivo y li-

bre. Esto es un banco de información en diferentes idiomas que incluye textos y oraciones monolingües y bilingües de tipología muy variada. Para construir este banco de información o corpus, se recopila de manera automática cualquier contenido útil que se encuentre en internet.

El corpus desarrollado con la colaboración de la empresa del PCUMH contiene actualmente subcorpus en un total de 75 idiomas, compuesto por textos sin traducir; y 18 pares de lenguas, compuestos por textos traducidos. El corpus multilingüe desarrollado por la empresa del PCUMH sirve para nutrir los conocidos como Large Language Models (LLMs). En concreto, estos son modelos desarrollados por Inteligencia Artificial y diseñados para comprender y generar lenguaje humano de manera avanzada. Uno de los ejemplos de LLMs más representativos sería ChatGPT. Para lograr su objetivo, es fundamental que este tipo de aplicaciones cuenten con acceso a corpus como el diseñado por la empresa Prompsit.