

Objetivo del proyecto

Investigación sobre Procesamiento Natural de Lenguaje usando modelos generativos pesados para la resolución de diversas tareas incluidas en la infraestructura de Clariah-es.

Participantes del proyecto

HiTZ: Centro Vasco de Tecnología de la Lengua, www.hitzeus.es/hitz

Universidad Nacional de Educación a Distancia - UNED, www.uned.es

Universidad Complutense de Madrid, www.ucm.es

Universitat d'Alacant, www.ua.es

Universidad de Jaen, www.ujaen.es

Universidade de Santiago de Compostela, www.usc.gal

Biblioteca Nacional de España, www.bne.es

Consejo Superior Investigaciones Científicas, www.csic.es

Barcelona Supercomputing Center, www.bsc.es

Universidad de Las Palmas de Gran Canaria, www10.ulpgc.es

Dialnet, <https://dialnet.unirioja.es>

SCAYLE, Supercomputación Castilla y León (España), www.scayle.es

Funciones de SCAYLE

SCAYLE ofrece la oportunidad de trabajar con grandes modelos de lenguaje de gran peso (>8B de parámetros) pudiendo explorar técnicas con gran coste computacional como *Supervised Fine-Tuning*, *Self-Improving strategy* y *Reinforcement Learning*. Gracias a las GPUs de gran potencia disponibles y a la infraestructura eficientemente interconectada de la partición Genoa disponen de la oportunidad de usar técnicas de paralelización (*DeepSpeed*, *3D parallelism*) y métodos de entrenamiento para modelos muy grandes de manera muy eficaz, pudiendo ejecutar procesos que en máquinas locales no sería posible. Además, gracias al *software* y compiladores disponibles en el servidor, pueden explorar técnicas y *software* novedosos (*Flash Attention*, APEX, NVIDIA NeMO) y escribir código de bajo nivel para poder personalizar y adaptar las librerías a sus necesidades.

Justificación del proyecto

Gracias a Clariah-es y a su investigador principal, German Rigau, pueden tener acceso a la computación disponible en SCAYLE, concretamente acceso al servidor Caléndula.

Ejecución: 2025 al 2026.

Financiación del proyecto

Red de Clariah-es.

Líder del proyecto

CLARIAH-ES es una infraestructura digital de investigación distribuida constituida inicialmente por 10 entidades de investigación españolas que están relacionadas, por su participación anterior y su interés actual, con las infraestructuras europeas de investigación.



CLARIAH-ES

Referencia: RED2024-154077-E