

[Inicio](#) > ALIA, la primera infraestructura pública, abierta y multilingüe de IA en Europa

ALIA, la primera infraestructura pública, abierta y multilingüe de IA en Europa

El proyecto, coordinado por el BSC, proporciona modelos de lenguaje abiertos y transparentes para fomentar el uso del castellano y lenguas cooficiales en el desarrollo y despliegue de la IA



LA INFRAESTRUCTURA PÚBLICA DE IA ABIERTA Y MULTILINGÜE



En este contexto nace ALIA-40B, el modelo fundacional multilingüe público más avanzado de Europa, entrenado en el supercomputador MareNostrum 5

La iniciativa cuenta con financiación 100% pública para dar servicio de interés público y democratizar el acceso a la IA de la ciudadanía, la administración pública, las universidades y las empresas

El presidente del Gobierno de España, Pedro Sánchez, ha anunciado el lanzamiento del proyecto ALIA, la primera infraestructura pública europea, abierta y multilingüe que, gracias a las capacidades únicas de supercomputación del Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS), refuerza la soberanía tecnológica de España y Europa en el desarrollo de una inteligencia artificial transparente, responsable y al servicio de las personas.

ALIA es una iniciativa pionera en la Unión Europea para proporcionar una infraestructura pública de recursos de IA y servicios tecnológicos innovadores, como modelos de lenguaje abiertos para fomentar el impulso del castellano y lenguas cooficiales -catalán y valenciano, euskera y gallego- en el desarrollo y despliegue de la IA en el mundo.

El proyecto, que se inició con el Plan de Tecnologías del Lenguaje en 2019, está coordinado por el BSC, con el impulso y liderazgo de la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA), y forma parte de la Estrategia de Inteligencia Artificial 2024 del Gobierno de España. Iniciativas como AINA, impulsada por la Generalitat de Cataluña, e ILENIA, impulsada por la SEDIA, han sentado las bases para la construcción de esta infraestructura pública de IA.

Supercomputación pública para hacer avanzar la IA

Se trata de un proyecto abierto que se distingue por la transparencia y la apertura para impulsar la innovación y la adopción de las tecnologías, garantizando la fiabilidad tecnológica y la inclusión social y económica. La familia de modelos ALIA cuenta con la verificación de la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA) y está alineada con los estándares de transparencia que establece el Reglamento de IA.

Esta iniciativa pionera cuenta con financiación 100% pública para dar servicio de interés público y democratizar el acceso a la IA de la ciudadanía, la administración pública, las universidades y las empresas.

“El proyecto ALIA representa un esfuerzo extraordinario por dotarnos de datos, modelos de lengua y recursos propios dentro del entorno competitivo de la inteligencia artificial. En su núcleo, ALIA trabaja con textos en más de 35 lenguas europeas, asegurando una representación del 20% para las lenguas del Estado, lo que le convierte en el sistema de IA que mejor refleja nuestra realidad lingüística y cultural”, ha afirmado Mateo Valero, director del BSC.

Un gran modelo de lenguaje entrenado en MareNostrum 5

El entrenamiento y despliegue de la IA generativa requiere de una enorme capacidad de procesamiento computacional. En el caso del entrenamiento de la familia de modelos ALIA, el procesamiento de varios miles de millones de palabras necesita del uso de miles de horas de MareNostrum 5, uno de los superordenadores más potentes del mundo, ubicado y gestionado por el BSC.

En este contexto, el presidente del Gobierno de España ha anunciado también la publicación de ALIA-40B, el modelo fundacional multilingüe público más avanzado de Europa con 40.000 millones de parámetros, que ha sido entrenado durante más de 8 meses en el MareNostrum 5 con 6,9 billones de tokens (palabras o fragmentos de palabras usadas en estos sistemas) en 35 lenguas europeas. Su versión final estará entrenada con hasta 9,2 billones de tokens.

“El modelo ALIA-40B, con 40.000 millones de parámetros, supone un salto cualitativo respecto a su predecesor y es el primer modelo soberano y público de esta magnitud desarrollado en Europa, capaz de generar recursos especializados en áreas de interés social y económico”, ha añadido Valero.

El corpus de entrenamiento del modelo ocupa 33 terabytes de memoria, lo que equivaldría a 17 millones de libros, o 4,5 millones de fotos en alta resolución, o 6,6 millones de canciones. Estas cifras suponen un importante salto cualitativo respecto de su modelo predecesor 7B, con 7.000 millones de parámetros, que fue todo un hito como primer modelo desarrollado desde cero en España.

AINA e ILENIA, los precedentes

El proyecto ALIA se inició con el Plan de Tecnologías del Lenguaje en 2019. Proyectos como AINA, impulsado por la Generalitat de Cataluña, e ILENIA impulsado por la SEDIA, han sentado las bases para la construcción de esta infraestructura pública de IA. En la Estrategia Nacional de Inteligencia Artificial 2024, la puesta en marcha del proyecto ALIA es uno de los pilares clave para la creación de esta infraestructura pública de IA en castellano y lenguas cooficiales. Además, ALIA está alineada con el programa de la Década Digital de la Unión Europea, que guía la transformación digital de Europa y su soberanía tecnológica.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 1 Jun 2025 - 07:58): <https://www.bsc.es/es/noticias/noticias-del-bsc/alia-la-primer-infraestructura-publica-abierta-y-multiling%C3%BCe-de-ia-en-europa>