

# ExPlain4Health,

## Extracción de contenidos sanitarios profesionales y no especializados mediante inteligencia artificial

### Objetivo del proyecto

Los principales objetivos de este proyecto son:

- *Crear recursos lingüísticos en español en dominio sanitario;*
- *Investigar el uso de grandes modelos de lenguaje en aplicaciones de procesamiento de información médica;*
- *Evaluar la calidad y de los contenidos generados por IA con expertos humanos.*

Web del proyecto: <http://clarmed.csic.es/explain4health/>

### Participantes del proyecto

Consejo Superior de Investigaciones Científicas (CSIC), [www.csic.es/es](http://www.csic.es/es)

Real Academia Nacional de Medicina de España (RANME), <https://ranm.es>

Universidad San Pablo CEU, <https://info.uspceu.com>

Fundación Rioja Salud, [www.fundacionriojasalud.org](http://www.fundacionriojasalud.org)

Hospital General Universitario Gregorio Marañón, [www.comunidad.madrid/hospital/gregoriomaranon/](http://www.comunidad.madrid/hospital/gregoriomaranon/)

Hospital Regional Universitario de Málaga, [www.hospitalregionaldemalaga.es](http://www.hospitalregionaldemalaga.es)

SCAYLE, Supercomputación Castilla y León (España), [www.scayle.es](http://www.scayle.es)

### Funciones de SCAYLE

Se ha usado la infraestructura Caléndula de SCAYLE para hacer ajuste fino, *fine-tuning*, de modelos *Whisper* en español (tamaño "small" y "medium"). Se han empleado 30 horas de audio y las correspondientes transcripciones extraídas de vídeos de tema médico en español, tomados de fuentes autorizadas y licencia *Creative Commons* (ej., Portal Clínic, Sociedad Española de Cardiología, etc.). Los modelos entrenados se destinan a la transcripción automática del habla en dominio médico, a fin de mejorar los procesos de automatización de la toma de datos clínicos, entre otras tareas. Los datos se distribuirán en el repositorio Zenodo (<https://zenodo.org/records/16729213>) y los modelos ajustados se distribuirán en la plataforma Hugging Face (<https://huggingface.co/>).

**Ejecución:** 2025 al 2027.

### Financiación del proyecto

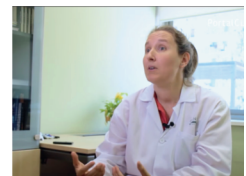
AEI/MICIU/10.13039/501100011033, convocatoria Generación del conocimiento 2024.

## Justificación del proyecto

El proyecto investiga cómo la inteligencia artificial generativa y el procesamiento del lenguaje natural (PLN) pueden mejorar el acceso a la información, simplificar y resumir contenidos médicos sin perder rigor. El objetivo es crear herramientas que reduzcan la sobrecarga de información para profesionales y faciliten la comprensión de contenidos para no expertos.

## Líder del proyecto

Leonardo Campillos Llanos, Instituto de Lengua, Literatura y Antropología (ILLA), Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC): Organismo público español, adscrito al Ministerio de Ciencia, Innovación y Universidades, la mayor institución pública dedicada a la investigación en España.



0.000000 4.3890209 la respuesta &mm concisa clara y breve es un no

4.3890209 9.5828622 pero siempre hay excepciones ¿no? es decir ya lo ha hablado también antes ¿no? la doctora Muñoz

9.5828622 19.6106252 es decir un no siempre y cuando no haya una clara sospecha de que un componente que lleve la vacuna de la covid

19.6106252 23.2498134 que se ha puesto de moda ahora sabéis el famoso polietilenglicol [politelengli'kol]



La imagen muestra fragmentos de vídeos del conjunto de datos (panel superior), una muestra de la transcripción y sus códigos de tiempo correspondientes (abajo a la izquierda) y un oscilograma o representación de onda del habla (abajo a la derecha).