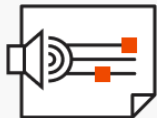


# HADA:

Investigación de  
Herramientas de  
Anotación de  
Datos  
Avanzadas

Mejoras en procesos de anotación en base a:

1. la **preparación** y **selección** de datos,
2. la **anotación automatizada** y
3. el **control** de la **calidad** de estos procesos



**SIGMA • COGNITION**

*El proyecto 2021/C005/00146323 está financiado por la UE Next Generation a través de la entidad pública empresarial adscrita al Ministerio de Economía y Transformación, RED.ES*

**SIGMA • COGNITION**

<https://sigmacognition.ai/>

Avenida de Manoteras 6,  
28050- Madrid.

Contacto: [esancho@sigmacognition.ai](mailto:esancho@sigmacognition.ai)



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

uc3m

Universidad  
Carlos III  
de Madrid



VICEPRESIDENCIA  
PRIMERA DEL GOBIERNO  
MINISTERIO  
DE ASUNTOS ECONÓMICOS  
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO  
DE DIGITALIZACIÓN  
E INTELIGENCIA ARTIFICIAL



Plan de Recuperación,  
Transformación y Resiliencia



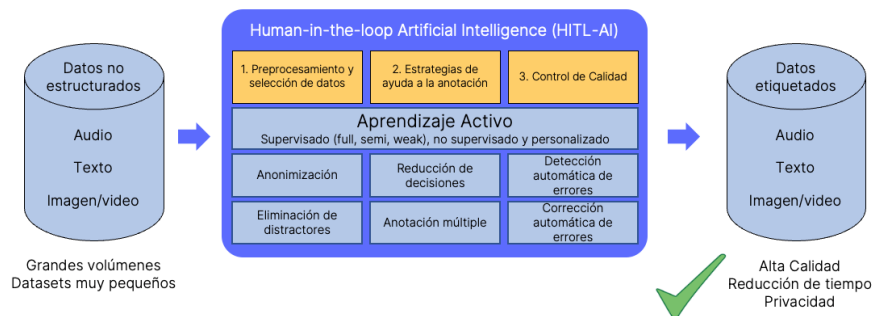
Financiado por  
la Unión Europea  
NextGenerationEU

red.es

# Objetivos de HADA

Investigación e implementación de **algoritmos de IA en el marco de la anotación de datos** para lograr a futuro **herramientas avanzadas de anotación** que permitan ayudar en la optimización de las tareas encaminadas a obtener datos de calidad, de una forma más rápida y precisa que las herramientas existentes actualmente.

- **Reducción del tiempo de anotación** mediante la **automatización de tareas repetitivas**, redirigiendo el esfuerzo humano a tareas de mayor valor añadido.
- **Selección de los datos de entrenamiento** que aporten más información a los **modelos** para **mejorar** su calidad de **funcionamiento** y **minimizar sesgos**.
- Desarrollar mecanismos que aseguren la **calidad de los datos**
- Asegurándonos de respetar las regulaciones sobre la **protección de datos**



**Aprendizaje Activo:** Investigación e implementación de modelos híbridos no supervisados y semi-supervisados para reducir la necesidad de grandes conjuntos de datos etiquetados.

**Reducción de decisiones:** modelo de IA que ayude en la reducción inteligente de opciones de etiquetado proporcionadas al anotador, tendiendo a problemas de clasificación binaria.

**Generación de datos:** Procedimiento para generar artificialmente datos mediante traducción, paráfrasis o generación de lenguaje

**Anonimización:** Aplicación de anonimización sobre los algoritmos que se empleen para selección de datos, ayuda a la anotación y control de la calidad.

**Detección automática de errores** en la anotación mediante técnicas de aprendizaje no supervisado.