

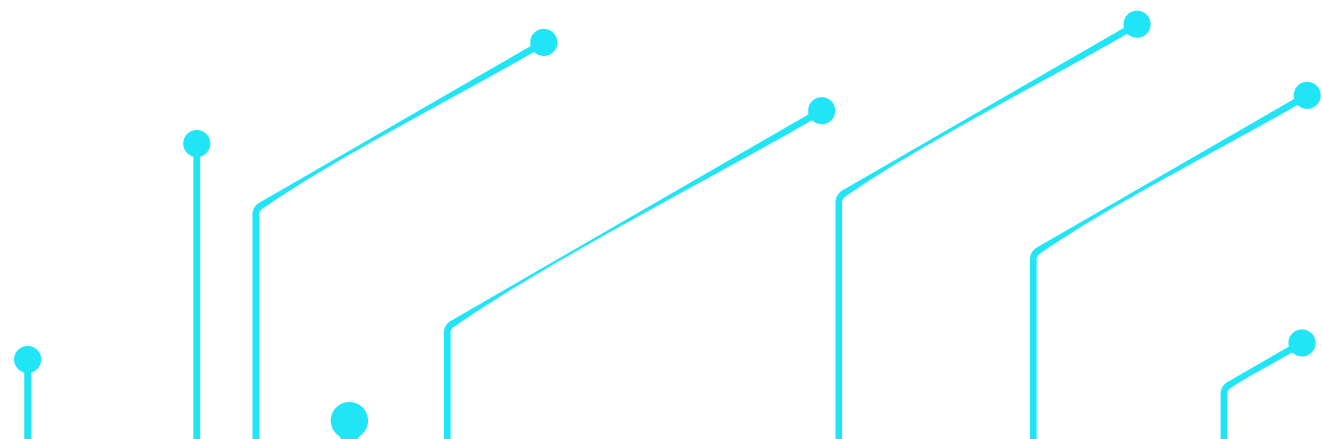
# Innovación SOAINT

DANIEL MEDINA



# Laboratorio 1

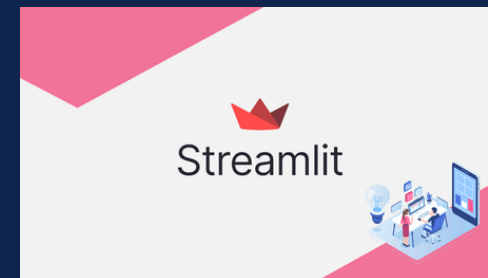
TRANSFORMACIÓN Y  
ANÁLISIS ENRIQUECIDO  
DE CV'S



# Importación de librerías

## STREAMLIT

interfaz de usuario



## PYPDF2

lectura de archivos pdf

## LANGCHAIN

procesamiento de lenguaje natural

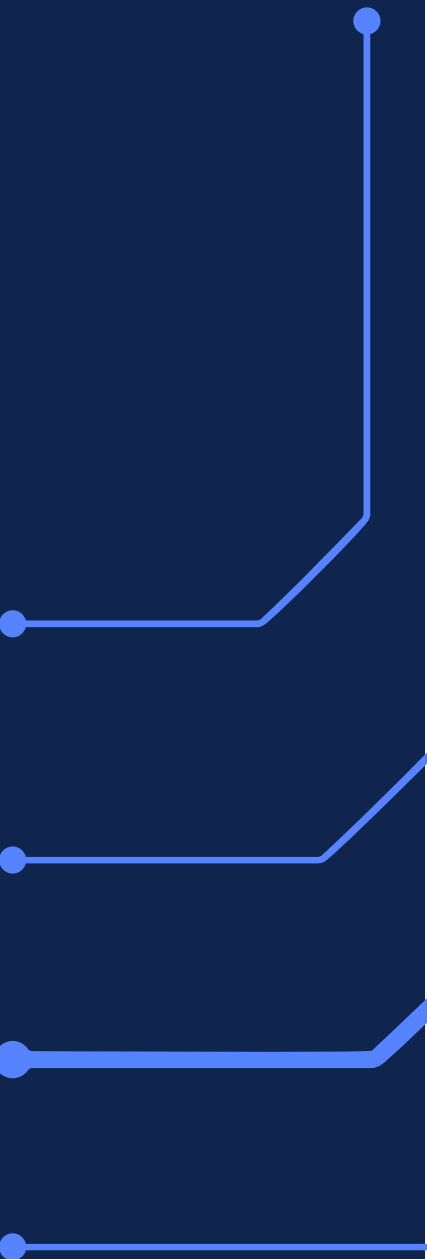
## FAISS

almacenamiento de vectores



# Proceso

- Se procesa el archivo PDF cargado.
- Extrae el texto del PDF utilizando PyPDF2.
- Luego, divide el texto en fragmentos más pequeños utilizando RecursiveCharacterTextSplitter.
- Crea embeddings (vectores numéricos) para cada fragmento utilizando el modelo sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 de Hugging Face.
- Almacena estos embeddings en un índice FAISS, que es un almacén de vectores eficiente para la búsqueda por similitud.



# Generación del CV en formato SOAINT

- Se crea un objeto DocxTemplate a partir de un archivo de plantilla Word (formato.docx).
- Se plantean una serie de preguntas al modelo de lenguaje GPT-4 de OpenAI utilizando la función similarity\_search de FAISS para buscar los fragmentos más relevantes del PDF. Las respuestas a estas preguntas se utilizan para rellenar campos específicos en la plantilla del CV, como nombre, nacionalidad, ubicación, resumen, formación, cursos, tecnologías y experiencia.



- Si el usuario ingresa una pregunta, se buscan los fragmentos más relevantes del PDF mediante `similarity_search` y se utiliza una función para generar una respuesta basada en esos fragmentos utilizando GPT-4. La respuesta se muestra en la interfaz de Streamlit.