

# **RECUPERACIÓN DE INFORMACIÓN AUMENTADA PARA GRANDES MODELOS DE LENGUAJE DE CÓDIGO ABIERTO**

UN CASO DE ESTUDIO EN AYUDAS PÚBLICAS DEL  
GOBIERNO DE ESPAÑA

Desiderio Martí Alcaraz

UOC - Máster en Ciencia de Datos (Área 2 - Aula 1)

Tutor: José Luis Iglesias Allones



AyudaMe.ai

# Motivación y justificación de interés

## Sobre mi

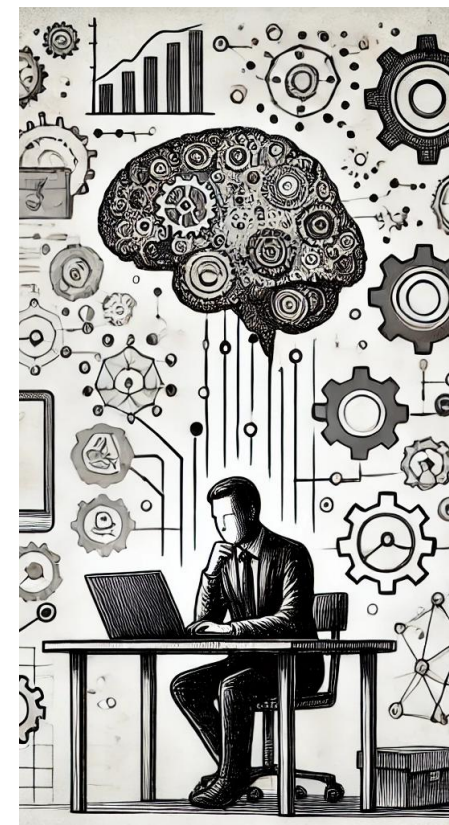
- Ingeniero de software especializado en intranets, portales documentales y plataformas colaborativas.
- Actualmente Responsable Especialista en Análisis de Datos y Reporting en Dir. Organización.

## Motivación del proyecto:

- Explorar la técnica RAG y grandes modelos del lenguaje para búsqueda de datos estructurados y no estructurados.

## Objetivo a largo plazo:

- Validar su potencial uso en búsquedas de sistemas documentales internos e intranets.



# Problema identificado



La información sobre ayudas y subvenciones públicas en España es compleja, voluminosa y poco accesible. Esto dificulta que ciudadanos y empresas encuentren oportunidades relevantes, generando barreras de entrada.

**Causas:** Presentación de datos no estructurada y búsqueda manual ineficiente.

**Impacto:** Pérdida de oportunidades y desigualdad en el acceso a recursos públicos clave para el desarrollo social y económico.

**Estado actual:**

- Los sistemas existentes son limitados: búsquedas tradicionales y filtrados básicos.
- Falta una solución que procese grandes volúmenes de datos, integre información y ofrezca respuestas personalizadas.

# Objetivos del proyecto



## **Objetivo general:**

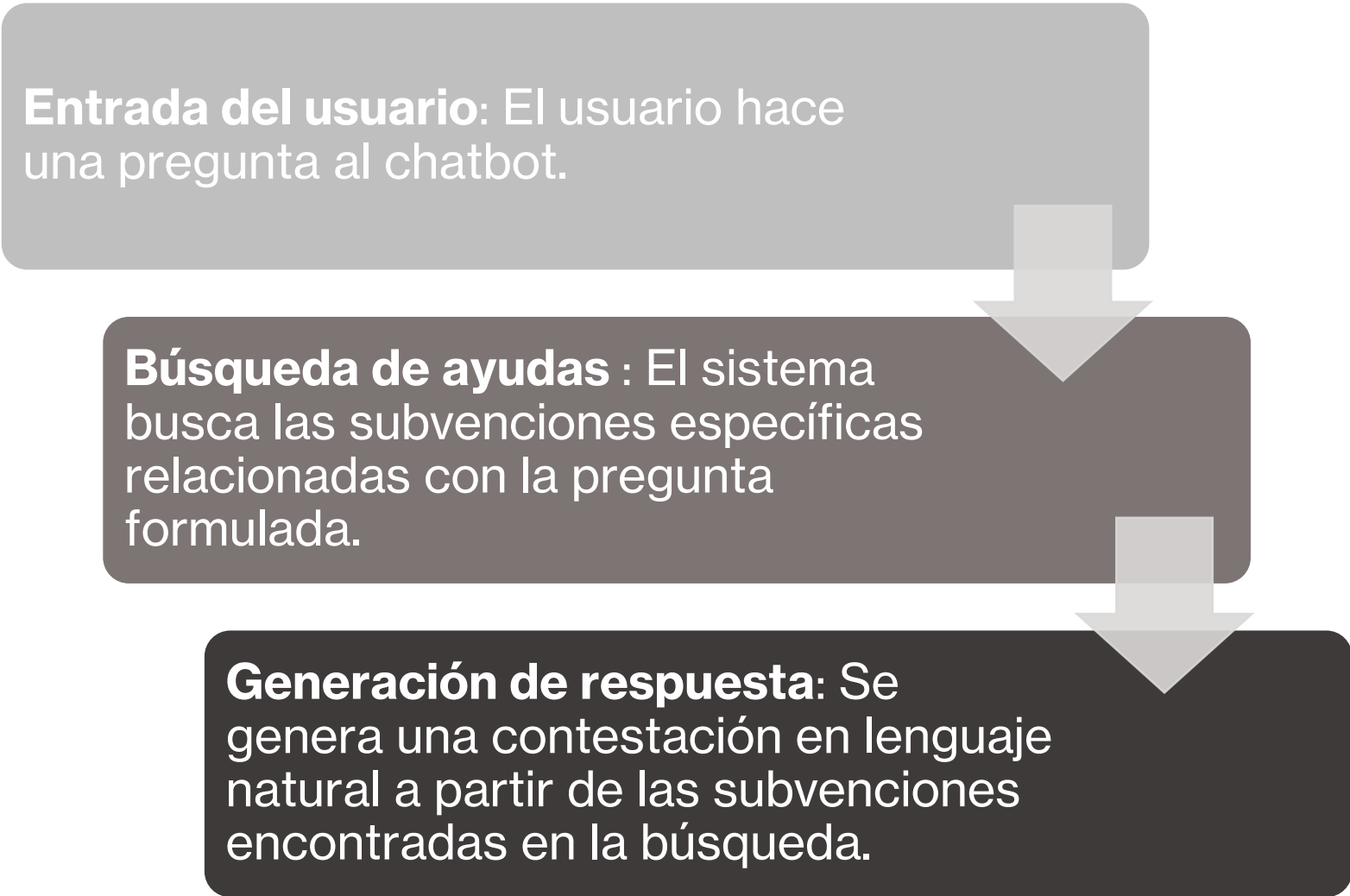
- Desarrollar un asistente virtual que facilite la búsqueda de información sobre ayudas y subvenciones públicas.

## **Objetivos específicos:**

- Extraer datos de ayudas de fuentes oficiales.
- Organizar información extraída.
- Crear un chatbot para responder a preguntas de los usuarios con respuestas en lenguaje natural que los usuarios puedan entender fácilmente.

# — ¿Cómo funciona?

**Entrada del usuario:** El usuario hace una pregunta al chatbot.



```
graph TD; A[Entrada del usuario: El usuario hace una pregunta al chatbot.] --> B[Búsqueda de ayudas : El sistema busca las subvenciones específicas relacionadas con la pregunta formulada.]; B --> C[Generación de respuesta: Se genera una contestación en lenguaje natural a partir de las subvenciones encontradas en la búsqueda.];
```

**Búsqueda de ayudas :** El sistema busca las subvenciones específicas relacionadas con la pregunta formulada.

**Generación de respuesta:** Se genera una contestación en lenguaje natural a partir de las subvenciones encontradas en la búsqueda.

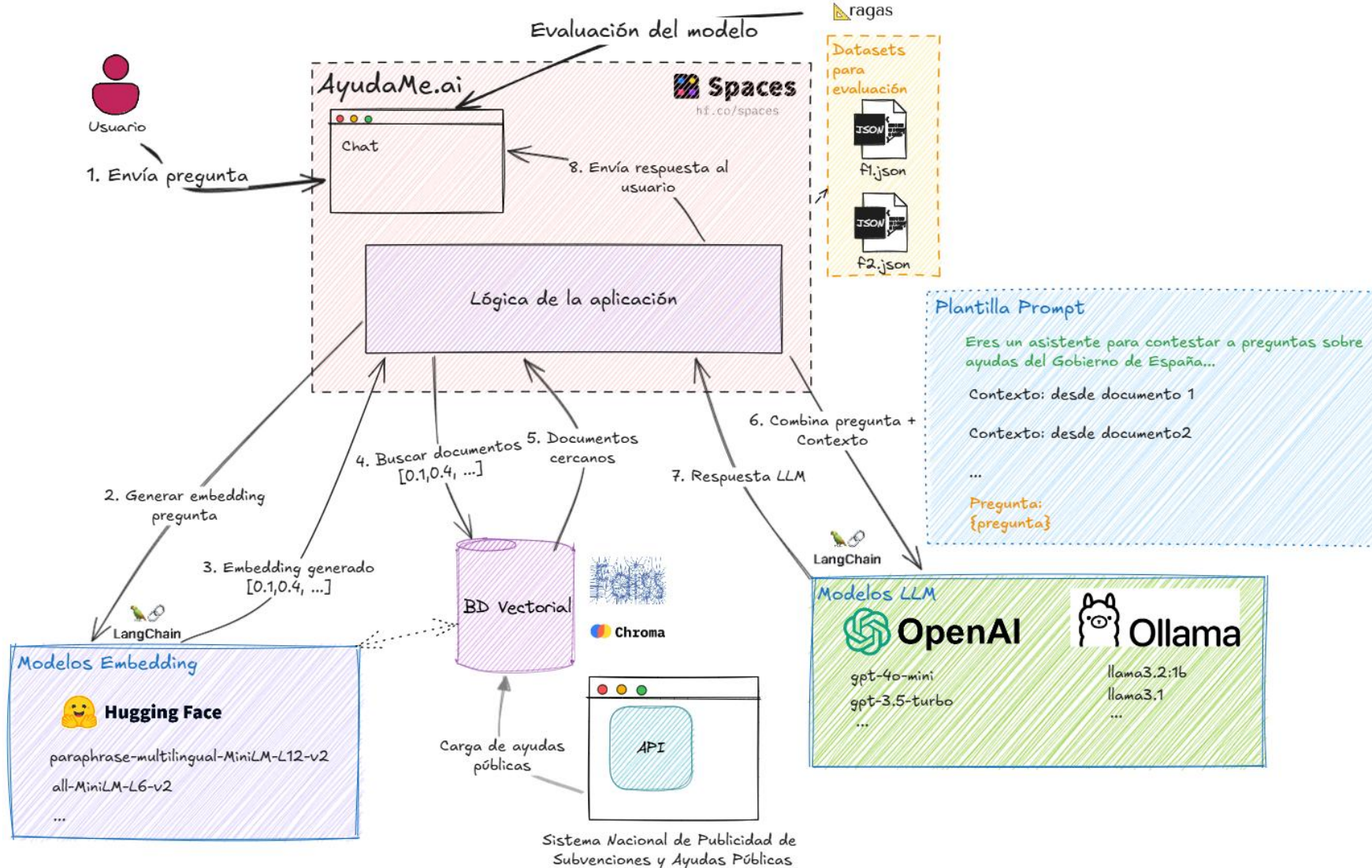
# Solución técnica

Implementación de un sistema basado en RAG (Retrieval-Augmented Generation) y LLMs.

- Extracción de ayudas mediante API REST o subida manual de documento PDF.
- Almacenamiento de ayudas
  - Modelo de Embedding (paraphrase-multilingual-MiniLM-L12-v2, all-MiniLM-L6-v2, )
  - BD Vectorial entre ChromaDB y FAISS (ambas Open Source).
- Generación de respuesta mediante un LLM, seleccionar entre:
  - Ollama (Open Source), modelo (Llama 3.2:1b, ...)
  - OPENAI (Con coste), modelo LLM (gpt-3.5-turbo, gpt-4o-mini, ...)
- Generación de preguntas sintéticas, seleccionar entre:
  - LLM (Ollama y OPENAI)
- Evaluación del sistema mediante framework RAGAS, parametrizable:
  - LLM (Ollama y OPENAI)



# Diagrama arquitectural



# Caso de uso real...



**Usuario:** ¿Qué tipo de ayuda ofrece el ayuntamiento de Pollença en 2025?



**Asistente:**

El Ayuntamiento de Pollença ofrece en 2025 una subvención y entrega dineraria sin contraprestación. Esta ayuda está destinada a asociaciones de índole social y tiene un presupuesto total de 80,000 euros.

Detalles de la convocatoria:

- Tipo de ayuda: Subvención y entrega dineraria sin contraprestación
- Convocatoria: Concesión de subvenciones a las asociaciones de índole social año 2025
- Tipos de beneficiarios: PYMEs y personas físicas que desarrollan actividad económica
- Sectores involucrados: Otros servicios
- Región de impacto: Mallorca (es532)
- Finalidad: Servicios sociales y promoción social
- Fecha de inicio de la solicitud: 1 de febrero de 2025
- Fecha de fin de la solicitud: 10 de febrero de 2025

Para más información, puedes consultar la convocatoria en el siguiente enlace:

[Convocatoria Ayuntamiento de Pollença 2025](<https://www.pap.hacienda.gob.es/bdnstrans/ge/es/convocatorias/806696>).



# — Evaluación del sistema

El sistema realiza una evaluación continua del desempeño del sistema, tanto de la búsqueda de ayudas como de la respuesta generada en lenguaje natural.

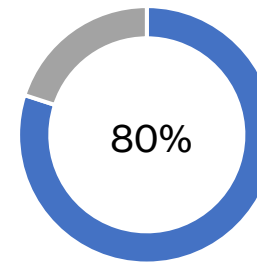
## ¿Qué datos usamos para la evaluación?

- Datos de búsquedas reales.
- Datos sintéticos:
  - Basados en plantillas de preguntas.
  - Generados por inteligencia artificial.



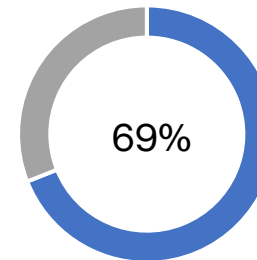
# Resultados obtenidos

- **Consistencia:** El modelo es muy consistente al generar respuestas similares al texto de referencia.
- **Alta Variabilidad:** En términos de relevancia y precisión del contexto, el desempeño es menos constante, mostrando comportamientos irregulares.
- **Desempeño Aceptable:** Las respuestas suelen ser fieles al contexto proporcionado, aunque hay casos específicos en los que podría mejorar.
- **Desempeño Deficiente:** Hay problemas en la recuperación del contexto y en la precisión de los hechos.



## Similitud semántica con pregunta

Similitud semántica con el texto de referencia



## Fidelidad al contexto

Fidelidad de las respuestas al contexto proporcionado

# Fortalezas y debilidades



## Fortalezas

El modelo destaca por generar respuestas consistentemente similares al texto de referencia. Algunas respuestas alcanzan niveles de relevancia y precisión completamente satisfactorios.



## Debilidades

Existen inconsistencias notables en la relevancia de las respuestas y la precisión.



## Áreas de Mejora

Es necesario mejorar la capacidad para recuperar contexto relevante y garantizar una mayor precisión en las respuestas.

# Conclusiones



## Logros alcanzados

- La solución ha logrado cumplir con los objetivos inicialmente previstos, principalmente con los relacionados con el modelo LLM y la recuperación de la información.
- Evaluación satisfactoria del sistema mediante el uso del framework RAGAS.
- Base sólida para continuar con la investigación identificando como área crítica la gestión de bases de datos vectoriales y la optimización de consultas.

- 
- El rendimiento de la base de datos vectorial no fue el esperado, a pesar de múltiples intentos de mejora: cambios de formato, modelos de embeddings, algoritmos de búsqueda y bases vectoriales distintas.
  - Implementación inicial fue planteada con herramientas Open Source (Ollama + ChromaDB / FAISS) en local, sin embargo, los problemas para encontrar infraestructura cloud obligaron a adaptar la solución al uso de OPENAI en HuggingFace Spaces.



## Desafíos encontrados

# Trabajos futuros

## Prioritarios

- Revisar desempeño de base de datos vectorial:
  - Explorar embeddings específicos en español.
  - Revisar parametrizaciones y otras bases de datos vectoriales.
- Agregar metadatos relevantes (geolocalización detallada, categorías específicas, etc.).
- Hacer comparativa de diferentes tecnologías que justifiquen el uso o no de tecnologías Open Source.

## Secundarios

- Mejorar usabilidad y desempeño del chatbot.
- Mejorar la interfaz de usuario.





# Demo

Web:

<https://huggingface.co/spaces/DesiMarti/TFMCienciaDatos>

Código Fuente:

[https://github.com/desimartiout/TFM\\_Desimarti](https://github.com/desimartiout/TFM_Desimarti)

