



Problema Sesión 02

Diseño y Estrategias en Producción
para Soluciones de IA

DESARROLLO DE UN SISTEMA RAG Y EXPERIMENTACIÓN SISTEMÁTICA

En esta práctica te retamos a que **desarrolles un chatbot** sencillo que recomiende películas a los usuarios. También queremos que realices **una experimentación sistemática** que te permita definir cuál es el mejor enfoque, de acuerdo con una serie de métricas.

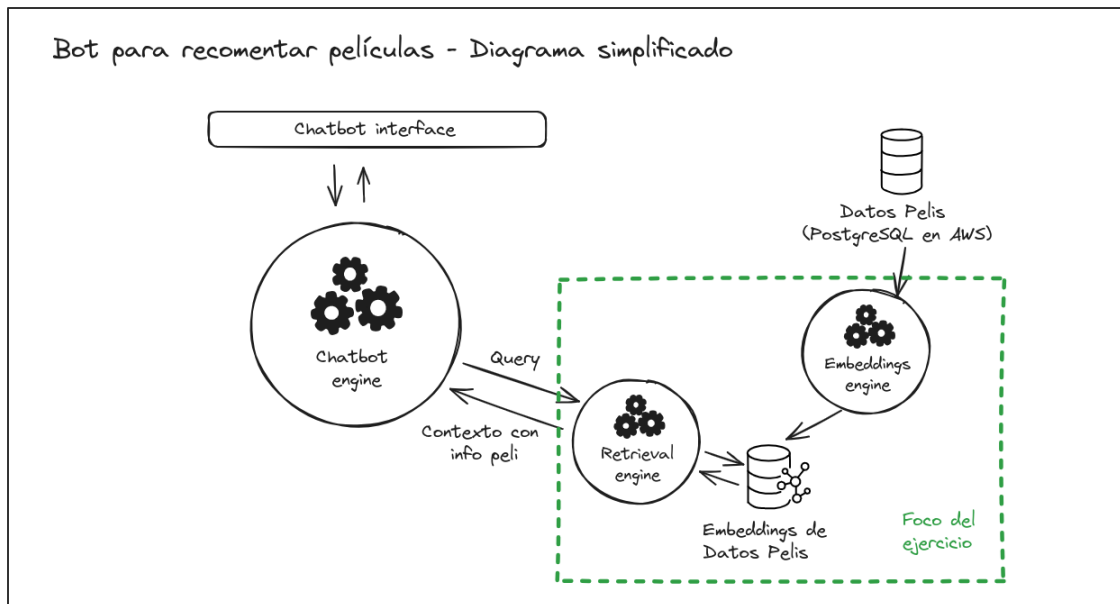
Objetivos de este ejercicio

El objetivo de este ejercicio es entender en detalle cómo funciona un sistema de RAG (*Retrieval Augmented Generation*) e ilustrar algunos de los aspectos comentados en el fastbook y en clase sobre la definición de métricas de evaluación y la experimentación sistemática en el desarrollo de soluciones IA.

Descripción de la actividad

La idea de esta práctica es que diseñes una versión muy simplificada de un chatbot en español que sea capaz de recomendar películas a los usuarios, utilizando las técnicas de RAG.

En concreto, queremos que te centres en **el módulo de *retrieval***, que es una de las partes más importantes de un sistema RAG (si no la que más). Por último, realiza varios experimentos para mejorar la métrica de evaluación que has definido.



Los pasos que debes seguir son los siguientes:

1. Volver a hacer un **fork** del repositorio con el que ya trabajaste en el lab anterior:

<https://github.com/RR-28023/bootcamp-ia-sprint-4>

Este paso es importante, porque el ejercicio no se publicará en el repo hasta las 00:05 h (CET) del día marcado para la apertura de este lab, así que no vale con el fork que has realizado en el primer lab.

2. A partir de ahí tendrás que seguir las instrucciones en el README. **Aviso importante:** por razones obvias, las credenciales para conectarse a la base de datos no se han incluido en el README del repo, pero aquí te las compartimos ahora:

```
RDS_USER=alumno_qualentum
RDS_PW=a78dysdg919z
```

No obstante, en la sesión en directo y la grabación de esta, el profesor-instructor explica cómo realizar el ejercicio. Cualquier duda que te surja, puedes trasladarla durante la clase o por los diferentes canales de comunicación de la Escuela.

Formato de entrega

El entregable debe ser una captura de pantalla del *dashboard* de MLFlow con tus experimentos, según se detalla en el README del repo.

Criterios de corrección

Tu ejercicio se dará por válido, si realizas correctamente los pasos marcados en este documento y adjuntas la evidencia de ello en el entregable.



[Qualentum.com](https://www.qualentum.com)