

Aplicaciones LLM

App básica: Evaluación de una app QA

Problema

- Queremos evaluar la calidad de una aplicación de preguntas y respuestas sobre un documento (QA App).
- Evaluar una aplicación LLM no es fácil, pues ante las mismas preguntas las respuestas pueden ser ligeramente distintas. No puede aplicarse un sistema de evaluación como el que utilizamos con las aplicaciones de software convencionales.

Solución

- Prepararemos una lista de preguntas de las que conocemos las respuestas y se las haremos a la QA App utilizando una RetrievalQA chain para hacer una evaluación manual.
- Después utilizaremos una QAEvalChain para que sea la propia App la que compruebe si sus respuestas coinciden con las respuestas de evaluación.

Proceso

- Cargar el documento de texto con un document loader.
- Dividir el documento en fragmentos con un text splitter.
- Convertir los fragmentos en embeddings con OpenAIEmbeddings.
- Cargar los embeddings en una base de datos vectorial FAISS.
- Crear una chain RetrievalQA para recuperar los datos incluyendo una input_key para identificar el prompt del usuario (la pregunta).
- Creamos un diccionario con las preguntas y respuestas de evaluación.
- Utilizamos la RetrievalQA chain para evaluar manualmente la App.
- Utilizamos una QAEvalChain chain para que la App se evalúe a sí misma.