



RAG Exam

Presentation

By Thabet Aljbreen, Nouga Aldawsari and Saud Abu Sarraah

Objective:

implement a Retrieval-Augmented Generation (RAG) system that uses a language model and a vector database to answer questions about traffic violations.

Problems we faced

Generate Embedding

```
from langchain.embeddings import HuggingFaceEmbeddings
embeddings = HuggingFaceEmbeddings(model_name="sentence-transformers/all-MiniLM-L6-v2")  
  
from langchain_huggingface import HuggingFaceEmbeddings  
  
embeddings = HuggingFaceEmbeddings(model="sentence-transformers/all-MiniLM-L6-v2")
```

Overlab

```
text_splitter = RecursiveCharacterTextSplitter(  
    chunk_size=250,  
    chunk_overlap=20,  
    length_function=len,  
)
```

Features for our Model

- The model answer you in English and Arabic
- answer you within the document you upload
- you can modify the prompt upon your desire

The Prompt and the answer

```
[ ] query = input("اسأل أي سؤال ؟")
```

اسأل أي سؤال ؟: ماهي الغرامه على القيادة بدون رخصه؟

```
[ ] response = query_rag(query)  
print(response)
```

.الغرامه على القيادة بدون رخصه هي 500 ريال إماراتي

DEMO with
our team
member





Thank You

For your time