Jupiter FAQ Bot - Project Deliverables

Project Overview

Title: Building a Human-Friendly FAQ Bot from Jupiter Help Centre
Objective: Create an intelligent, conversational bot to answer Jupiter users' queries based on FAQ data scraped from the website.

---

1. Methodology and Architecture

Data Collection

- Scraped internal links from https://jupiter.money
- Filtered pages containing `<h1>` with text like "Frequently Asked Questions"
- Extracted Q&A pairs using:
  - `div.faq-header` and `div.faq-answer`
  - `h3` headings with following `<p>` as fallback

Preprocessing and Cleaning

- Removed HTML tags, excessive whitespace
- Deduplicated questions using:
  - TF-IDF vectorization
  - Cosine similarity thresholding (0.9)

Semantic Retrieval Pipeline

- Embedded questions using SentenceTransformer (`all-MiniLM-L6-v2`)
- Built FAISS index for fast vector similarity search
- Retrieved top match for user query

LLM-based Answer Generation

- Used **Mistral-7B-Instruct-v0.2** for rephrasing matched FAQ answers in a friendly tone

- Prompt Template:

```
[INST] You are a helpful support assistant for Jupiter. Rephrase the
following answer to be friendly:
Question: ...
Answer: ...
Friendly Response: [/INST]
```

Tech Stack

Component	Tool/Library
Web Scraping	BeautifulSoup, requests
Semantic Search	SentenceTransformers + FAISS
LLMs	Mistral-7B-Instruct
Model Framework	Hugging Face Transformers

2. Evaluation of Semantic Similarity & Relevance

Evaluation Criteria

- **Relevance**: Does the retrieved FAQ match the intent of the query?
- **Fluency**: Is the rephrased answer clear and human-friendly?
- **Fallback Handling**: If no match found, does the bot respond gracefully?

Sample Evaluation

User Query	Retrieved Question	Score
How can I pay my water bill?	Can I pay my bills using Jupiter?	0.92
What happens if EMI bounces?	What happens if I miss a payment?	0.89

- Relevance score calculated using cosine similarity of embedding vectors

3. Notebook / Application

File Provided:

- `jupiter_faq_bot.ipynb`: Contains all stages from scraping to LLM response
- Supports:
  - Offline GPT-2 generation
  - Mistral 7B chat (if GPU available)
  - Streamlit/CLI interface for user input