AI Store Assistant Project Report

Cuong Dang

2024-10-15

1. Project Overview and Objectives

This project aims to create an AI Store Assistant for a tech company using GPT-4o-mini LLM model.

2. Methodology

2.1 Compiling codes

Combine the codes given in the End-to-End-AI-App-with-Evaluation.ipynb. The main objective was to understand what each function do and how they work together.

2.2 Key Functions and Their Purposes

- `process_user_message`: Process the user inquiries and moderate it for the appropriateness. Once evaluated, then generate a suitable response from the products.json along with helper functions from utils.py. If the evaluated response is appropriate, then return it to the user.
- `log_messages`: log the timestamp, user's inquiry, AI response, and the metadata for each inquiries.
- `collect_messages_en`: Create a chain prompt system so that the AI response can have a context from the previous conversations.

3. Challenges Faced and Solutions Implemented

- **Challenge**: Unable to implement the `collect_messages_en`
  - **Solution**: Create a simpler version and keep building off of it until it has all the functions similar to the provided function in the End-to-End-AI-App-with-Evaluation.ipynb.
- **Challenge**: Logging only the latest message
  - **Solution**: Build a separate function that reads the previous logs and then append to it, and then rewrite the new log.
- **Challenge**: Difficulty in applying gradio
  - **Solution**: Refer back to previous projects and understand how the demo works.
- **Challenge**: Deploying on Hugging Face
  - **Unresolved**: Not sure how to use a LLM model from Hugging Face.

4. Potential Improvements and Future Work

- Learn how to deploy on Hugging Face
- Start the project earlier to have enough time to do research on Hugging Face