

Athina AI - Internship application

Documentation

1. How you constructed your dataset

I created the dataset using OpenAI by first processing the PDF document and then instructing ChatGPT to generate 100 query-response pairs covering all aspects of the document, including tables and different sections across various pages. After receiving the responses, I sorted and reviewed the questions to select approximately 30 high-quality query-answer pairs for the final dataset. This method ensured comprehensive coverage of the document and maintained a high standard of relevance and accuracy.

2. How and why you chose these evaluation metrics

I selected these evaluation metrics for their clear alignment with the primary objectives of the project and their ability to comprehensively assess the chatbot's performance:

Accuracy-focused Criteria: The `'accuracy_criteria'` dictionary enables precise evaluation of the chatbot's responses on a scale from 1 to 10. This ensures a nuanced assessment of the correctness and factual accuracy of the provided answers, directly addressing the delivering accurate information to users.

Relevance: Relevance is implicitly measured by the requirement that the answers align closely with the reference context. This ensures that the chatbot's responses are directly related to the user's query, providing information that is relevant and useful.

Completeness: Completeness is also implicitly measured by the requirement for the answers to align closely with the reference context. This ensures that the chatbot's responses cover all relevant aspects of the user's query, providing a comprehensive answer.

These evaluation metrics were specifically chosen to address the primary goals of the project, which prioritize delivering accurate, relevant, and complete information to users, while also ensuring the chatbot's responses are consistent and factually accurate across different queries and contexts.

3. What did you try to improve accuracy?

- **Handling Table Data:**

- Initially, I tried to extract tables directly using the `fitz` library, but I encountered issues where the term "Unlimited" extended over multiple columns, and only one column was identified correctly while the others were left blank.
- To improve this, I used `fitz` to find the table area within the PDF, converted that section to a PNG image, and then utilized OpenAI Vision to create JSON objects from the image tables. This approach significantly improved the accuracy of the information extracted from the tables.

- **Text Chunking:**

- Initially, I chunked the text into 500-character segments. However, this method resulted in fragmented table content and incomplete sentences.
- To address this, I kept tables as continuous chunks for clarity. For the rest of the sections, I used a combination of 500-character chunks followed by SpaCy NLP to ensure complete sentences were maintained within each chunk. This method preserved the integrity of the information and improved the coherence of the text chunks.

- **Cosine Similarity for Relevance:**

- After generating embeddings for the text chunks, I applied cosine similarity between the text chunks and the query vector to identify the top 10 most relevant chunks.
- These relevant chunks were then used as context for the OpenAI model, ensuring that the responses were contextually accurate and relevant.

- **Prompt Template:**

- I used a clear and well-defined prompt template to guide the OpenAI model, ensuring that it provided relevant answers without hallucinations. The prompt was designed to focus on extracting accurate and pertinent information from the given context.

These steps collectively improved the accuracy and quality of the responses generated by the chatbot, ensuring comprehensive coverage of the document's content while maintaining a high standard of relevance and accuracy.