

Blue Voice ML Take Home

Document Chatbot from Document Path

Project Objective

Create a Python script/file that when provided the path of a PDF document will start Q&A regarding the document. Context and history is not necessary, it can just support one-off questions.

Features

1. **Select a document:** Provide a path to the document as the first argument. The document can be up to 1000 pages long.
2. **Q&A:** Using LLM, RAG, or any other architecture, create a chatbot that will answer questions regarding the content of that document.
3. **Error Handling:** Display errors gracefully (e.g., file not found error etc.).
4. **Optional (Bonus Features):**
 - Provide the most relevant page of the document where the answer was found.
 - Connect it to a web/browser interface for a file uploader.
 - Add a UX for the chatbot.

Technical Requirements

- **Language:** Python
- **Invocation:** Please see the answer below.
- **Exit:** When the user types “exit” the program quits.
- **NOTE:** We will keep all third-party API keys confidential and will not use them outside of testing your code.

```
$ python chatbot.py ~/Desktop/document.pdf
Ingested document.pdf. Please ask your questions.
# What is the town that Nancy grows up in?
>>> Nancy grew up in Los Angeles.
# exit
```

Evaluation Criteria

1. **Code Quality**
 - Is the code well-organized and modular?
 - Proper use of version control.

2. Functionality

- Does it meet all the feature requirements?

3. Clarity

- Are the names of variables/functions/classes self-explanatory?

4. Algorithms and Approach

- Preprocessing algorithms for the PDF document.
- Chunking strategy for analyzing the document.
- Model chosen and why.

5. Documentation

- Is the project and codebase well-documented?
- Are there clear instructions to run the project?

6. Bonus

- Implementation of any additional features or considerations.

Submission Guidelines

- Host your code on a public repository like GitHub.
- Provide a README.md file with instructions to set up and run the project.
- Include any relevant documentation about your design choices and any libraries you decided to use.
- Once done, email the repository link for evaluation.

Good luck!