Simpplr DS Assignment

Policy QA system

As part of the 2nd round of evaluation for the position of Data Scientist at Simpplr, you have been given a task to build a QA system to answer policy questions. Simpplr being an employee experience company, we are using GenAI to make the day to day lives of employees easier, part of it helping employees quickly get answers to policy related questions like - “How many personal leaves do I have ?”, “How do I manage my ESOPs?”, etc.

For the given task, you have been provided with a few policy documents you can find in the pdfs folder.

Your task is to develop a system which can strictly answer any question that the user might have around policies available in any or all documents shared with you.

You need to -

* Come up with 2-3 different approaches to solve the problem
* Develop an evaluation criteria to evaluate the efficiency of your solution
* Build an a Rest API, which takes user\_query as input and returns response and sources , with response being the answer to the user query and sources being the chunks from the policy documents based on which the response is formulated.
* Containerize the Service, such that one can build and run the application using docker build and docker run commands.
* Create a README file to explain how to setup the service and file structure
* Write a Document explaining the various approaches you experimented with, your thought process on why you decided to try them out. Pros and cons of each solution and a table explaining the evaluation.
* Code should be of production quality with proper test cases defined

You can -

* Use Python to develop your solution
* Use any LLM model of your choice
* Use any type of document storage(Vector Store, keyword store, knowledge graphs, if using vector store please use elasticSearch )
* Use any python package of your choice to build the solution ( explain why you choose this package in documentation )

OPENAI\_KEY : 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

Please submit your solution in a zipped folder, or a private git repository.

Please feel free to reachout, in case of any confusions regarding the assignment