

## NLP Assignment: 5

### # Explanation for creation of Application



Follow. are main features for creation of application

#### 1) Understanding Requirement:-

Determine the types of Docs. & summarization goals

2) choosing Model :- Decide on an LLM like GPT-4 Turbo which are capable of handling large context windows and complex summarization tasks.

3) Environment Setup :- a) Install necessary python libraries such as 'langchain - openai' & 'tiktoken'; b) Set environment variable like 'OPENAI\_API\_KEY'

4) Document loading :- Use a 'PYPDF' loader or similar to load the documents that need summarization

#### 5) Summarization Approach :-

1) Stuff method :- Use a 'stuff doc chain' to combine all docs. into a single prompt and summarise them together

2) MAP :- Reduce Method :-

first map each doc. to an individual summary then reduce these summaries to a final comprehensive summary using 'MAP REDUCE DOC - CHAIN'.

## 6) Implementation:-

- Define prompt templates for mapping and reducing steps
- Implement the LLM chains with specific prompts for each step/stage
- combine the map and reduce chain for a final output

## 7) Execution & Testing:-

- split docs. If they are too long using tools like 'Character Text Splitter'.

Run the summarisation chain & evaluate the outputs for accuracy & completeness.

## 8) Refinement & Scaling:-

- Depending on the results refine the prompt & adjust the model parameters.
- consider implementing 'Refine Doc-chain' for iterative summarisation to enhance detail & accuracy.

## 9) Deployment:-

- Ensure the environment can scale especially if processing large volume of docs.
- Implement monitoring and logging to handle failures and ensure performance

## 10) Feedback loop:-

Collect user feedback on the summarise to continuously improve the summarization logic & ~~model~~ model finetuning.