PDF Summarizer

Technologies and Implementation

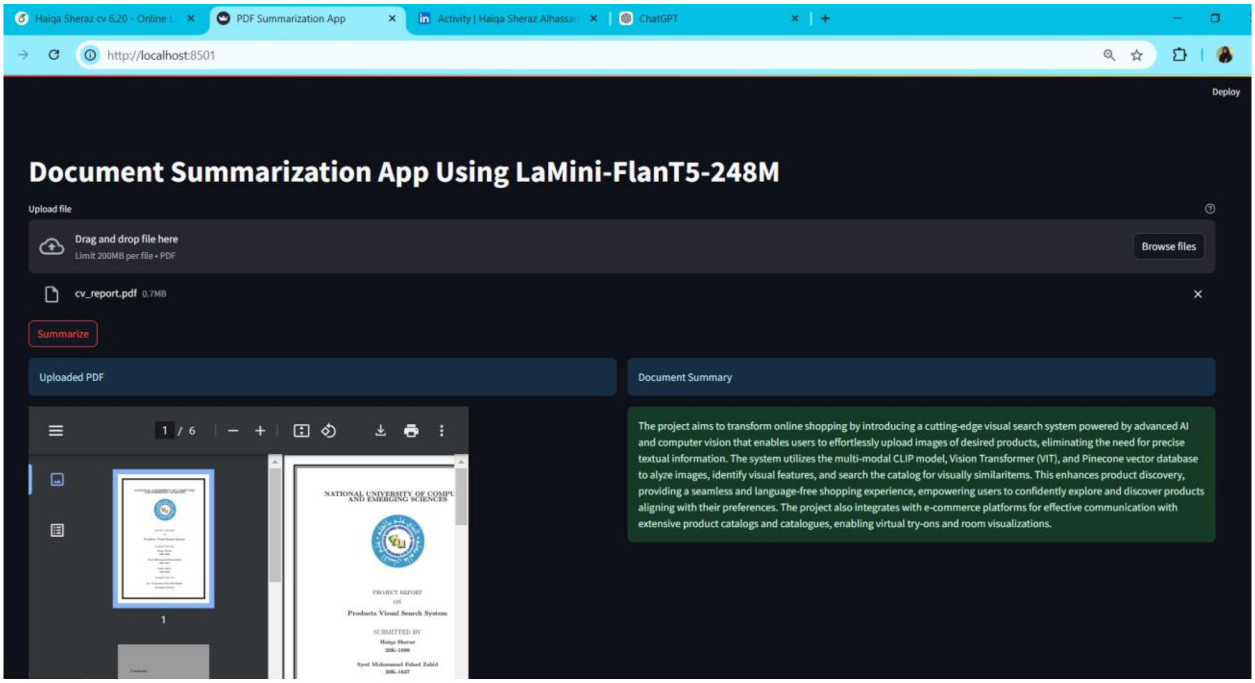
- Utilized the **LaMini-FlanT5-248M** model to implement a summarization pipeline for generating concise summaries of textual content.
- Integrated the **PyPDFLoader** library to process PDF files, extract text from up to 8 pages, and split the text into manageable chunks for efficient processing.
- The text chunks were passed to the summarization pipeline of the model to produce a coherent summary.
- Designed a **Streamlit application** to provide a user-friendly interface for uploading PDFs, viewing their content, and generating text summaries.

Limitations

1. **Restricted to Short PDFs**
   - The application is limited to processing PDFs of only a few pages. This is due to the `llm_pipeline` function, which aggregates all text chunks into a single string (`final_texts`). If the PDF contains a large amount of text, the combined content may exceed the model's input token limit, leading to truncation or errors during summarization.
2. **Inability to Handle Visual Elements**
   - The current implementation can only process text extracted from PDFs. This limitation arises because the model (`LaMini-FlanT5-248M`) and the `PyPDFLoader` library do not support visual or non-textual content such as images, graphs, or charts.

Improvements / Future Work

1. **Support for Large PDFs**
   - To handle larger PDFs, the text aggregation logic in the `llm_pipeline` function can be enhanced. Instead of concatenating all text into a single string, summaries could be generated for smaller batches of chunks and then combined into a final summary. This would allow processing of PDFs with more pages while adhering to the model's token limit.
2. **Incorporate Multimodal Processing**
   - Adopting a multimodal model like **GPT-3.5** or **GPT-4** would enable the application to process both textual and visual content, such as images or diagrams. These models can handle diverse input types and provide richer summaries that incorporate visual information. To achieve this, integration with APIs like OpenAI’s multimodal capabilities or tools like Azure Vision API could be explored.



Haiqa Sheraz cv 6.20 - Online L
PDF Summarization App
Activity | Haiqa Sheraz Alhassa
ChatGPT
http://localhost:8501
Deploy

Document Summarization App Using LaMini-FlanT5-248M

Upload file

Document Summary

The project aims to transform online shopping by introducing a cutting-edge visual search system powered by advanced AI and computer vision that enables users to effortlessly upload images of desired products, eliminating the need for precise textual information. The system utilizes the multi-modal CLIP model, Vision Transformer (VIT), and Pinecone vector database to alyze images, identify visual features, and search the catalog for visually similaritems. This enhances product discovery, providing a seamless and language-free shopping experience, empowering users to confidently explore and discover products aligning with their preferences. The project also integrates with e-commerce platforms for effective communication with extensive product catalogs and catalogues, enabling virtual try-ons and room visualizations.