**Sahil Wani**

SUID: 466848553

Apr 30th report

In this week’s report I started working on building an agent for summarizing the document user uploads. The agent should be able to read the text from uploaded pdf file or even the given URL and generate the summary using LLM. I am using cohere as my LLM to generate the summary and streamlit to create UI for chatbot.

**Key Technologies & Libraries Used:**

Streamlit:

Used for building the interactive user interface. Enables real-time inputs (file upload, text fields, buttons) and chat display.

Cohere:

Used to generate natural language summaries from raw text. We use the generate() endpoint to craft controlled summaries.

PyMuPDF (fitz):

Used to extract plain text content from uploaded PDF files.

Requests + BeautifulSoup:

Used to fetch and clean the content of web pages entered via URL.

**Workflow of chatbot I am currently working on will be:**

**UI setup:**  
The Streamlit interface collects:

* A PDF file (via st.file\_uploader)
* Or a URL (via st.text\_input)
* A summary length option: "Small" or "Large" (via st.radio)
* A trigger button: "Generate Summary"

**Text Extraction**  
Based on the user input:

* If a PDF is uploaded, I can define a function for extracting text input from pdf using **fitz**
* If a URL is entered, I can use BeautifulSoup to read the text from url.

**Summarization**  
The extracted raw text is then passed to cohere API using a function call where summary will be generated based on the length selected by the user for summary.

A prompt is dynamically crafted to instruct Cohere to summarize the content at the desired detail level. Max tokens are set based on the selected length:  
Small → 700 tokens, Large → 2000 tokens.

**Output Display**  
The summary is appended to a session-based conversation history and shown in a chat-like format using Streamlit UI.

Current Problems:

Currently I am working on issues related with connection of cohere API to my chatbot which I have used multiple times before for my chatbot.