**NUS Orbital 2024 - Milestone 2**

**Team Name:**KAYO

**Proposed Level of Achievement:**

Apollo 11

**Motivation:**

In academic settings, particularly during exam preparation or while studying textbooks, the need for timely and accurate clarification of doubts is paramount. However, traditional

methods of seeking assistance, such as sending entire PDFs or lengthy queries, often prove

inefficient and time-consuming. Recognizing this challenge, our project aims to streamline the process by developing a platform for sending concise prompts, reducing the cost of using AI based models like ChatGPT and generating accurate responses swiftly.

**Aim:**

We want to develop a web based platform that can make AI models like ChatGPT more

accurate and prevent any hallucinations due to inadequate or irrelevant data. The system will

process the document, dividing it into various chunks and embedding each chunk. The most

relevant chunks related to the query will be identified and a prompt will be sent to AI models

to further process and generate data.

This is so users can engage in educational research, data analysis and a variety of activities

without worrying about hallucinations and wrong outputs to a great extent from models like

ChatGPT.

**User Stories:**

1. As a researcher, I want to efficiently extract and accurately obtain information from my research papers, so that I can focus more on analysis and less on data retrieval.
2. As a student, I want precise and reliable answers to my questions from my textbooks and academic materials, so that I can better understand the concepts and prepare effectively for exams.
3. As an accountant, I want to accurately analyze vast databases to extract relevant information, so that I can make informed financial decisions and reports.
4. As a legal professional, I want to accurately retrieve relevant information from legal documents and case files, so that I can prepare stronger legal arguments and case strategies.

**Scope of Project:**

The KAYO web application aims to enhance the accuracy and reliability of AI models like ChatGPT in processing and extracting relevant information from documents. The platform is designed to assist users in academic, professional, and research settings by providing precise and reliable answers to their queries. This is achieved by chunking documents, embedding the chunks, and identifying the most relevant ones to send as prompts to AI models, thereby minimizing hallucinations and irrelevant outputs.

**Features completed:**

1. Core Processing Component:
   1. Query Handling: Develop a system for users to input concise prompts or queries.
   2. Document Chunking: Implement functionality to divide uploaded documents into manageable chunks.
   3. Embedding: Integrate embedding models to convert document chunks into vectors.
   4. Relevance Identification: Create algorithms to identify the most relevant document chunks based on user queries.
   5. Prompt Generation: Formulate the identified chunks into a coherent prompt for the AI model.
2. Basic User Interface:
   1. Query Submission Form: Design a form for users to submit their queries.
   2. Document Upload: Allow users to upload documents in supported formats.
   3. Display Results: Show the AI model’s responses in a user-friendly manner.

**Features yet to be completed:**

1. Improvement on the interface and refinement of the output.
2. Extended Data Format Support:
   1. Format Handling: Enable the platform to support multiple document formats (PDF, DOCX, TXT, etc.).
3. Custom Modes:
   1. Research Mode: Create a custom mode tailored for researchers, optimizing the retrieval process for academic papers and research materials.
   2. Student Mode: Develop a custom mode for students, focusing on extracting relevant information from textbooks and academic resources.
   3. Professional Modes: Implement custom modes for accountants and legal professionals to enhance data retrieval from financial documents and legal case files, respectively.
4. Feedback and Improvement Mechanism:
   1. User Feedback: Implement a feedback system for users to rate the accuracy and relevance of the responses.

**Tech Stack**

1. React
2. Node.js
3. Flask
4. Python
5. Langchain
6. HuggingFace
7. FAISS
8. Google Gemini API

**How are we different from similar platforms?**

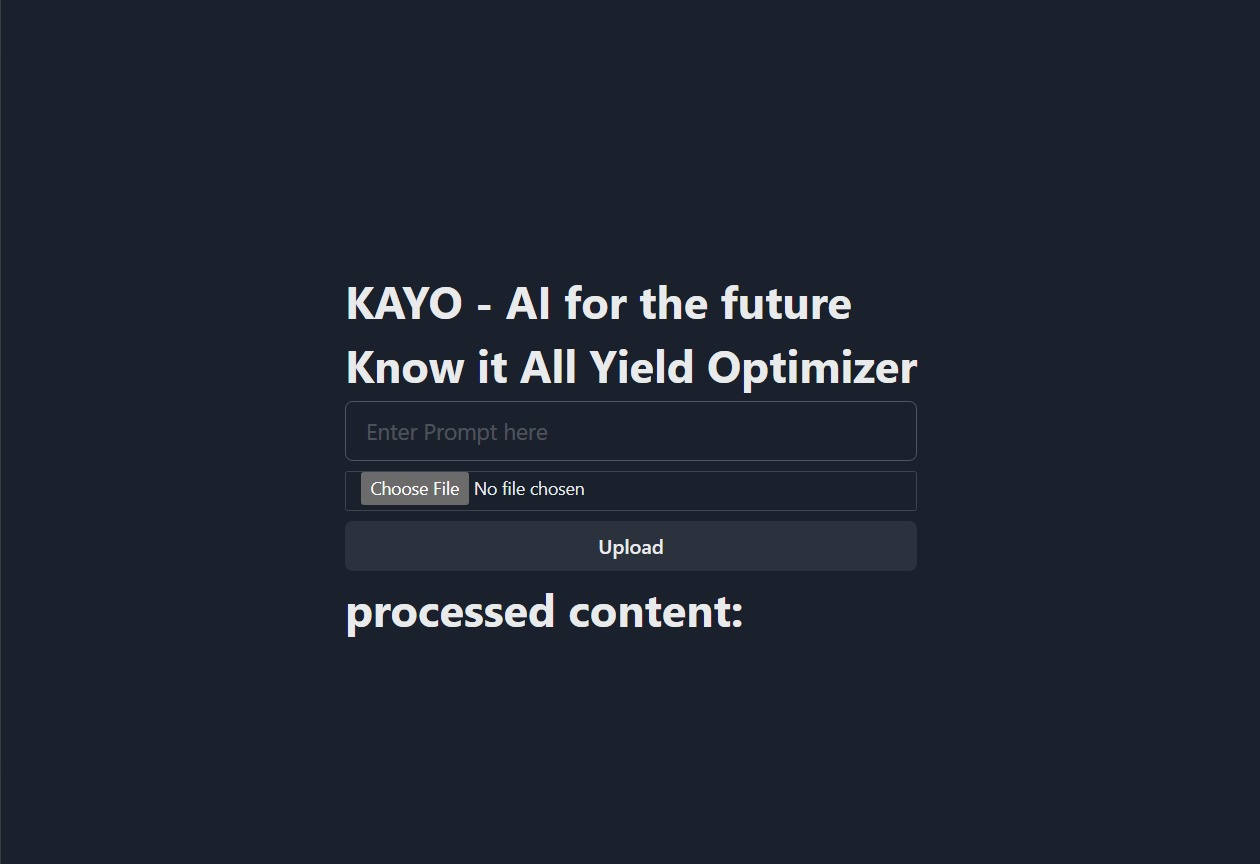
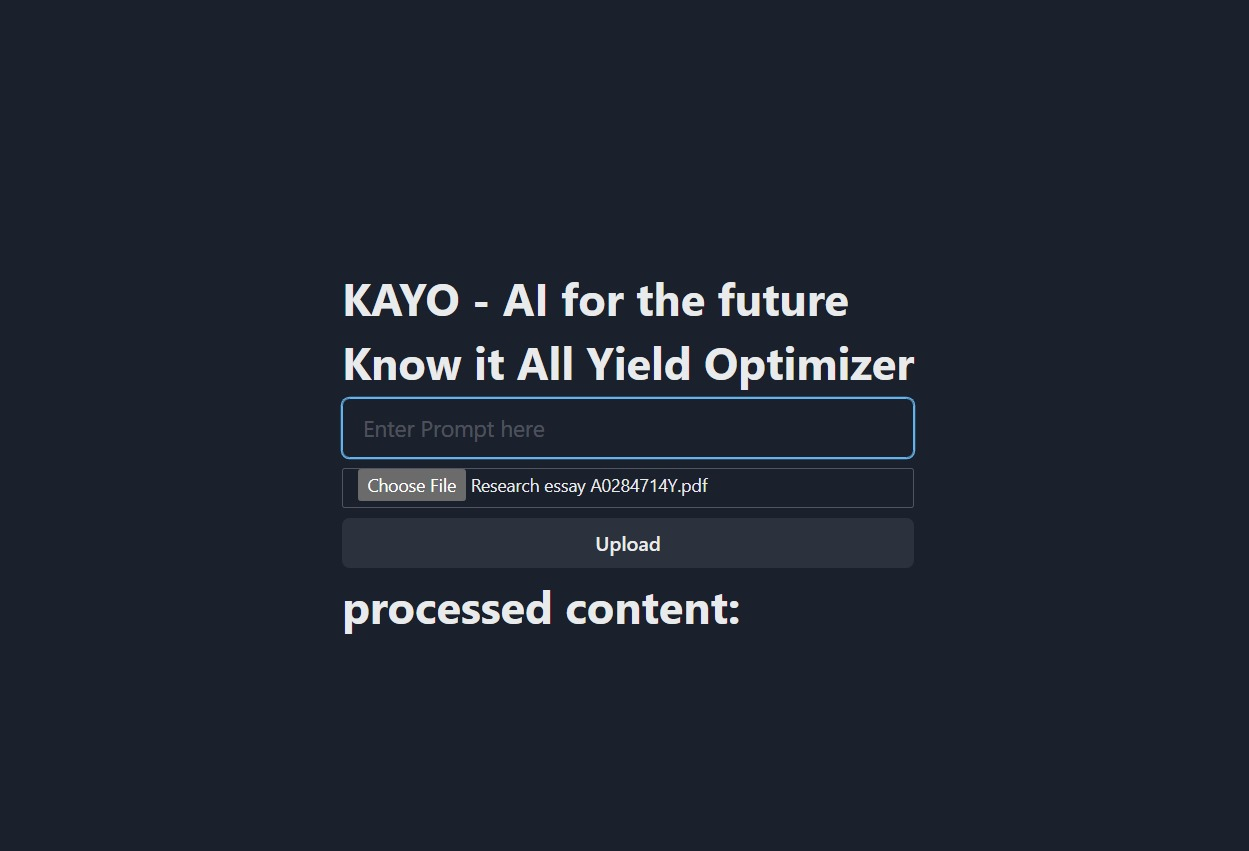
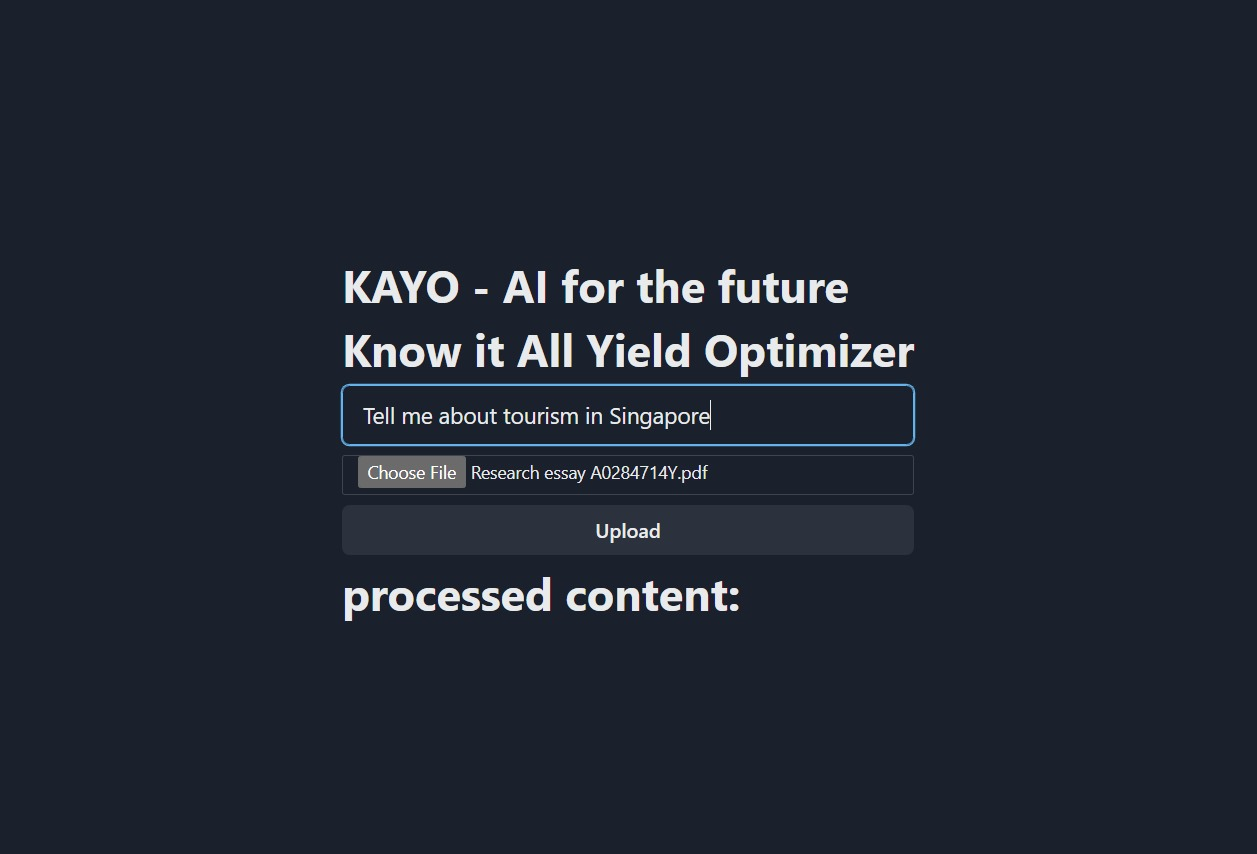
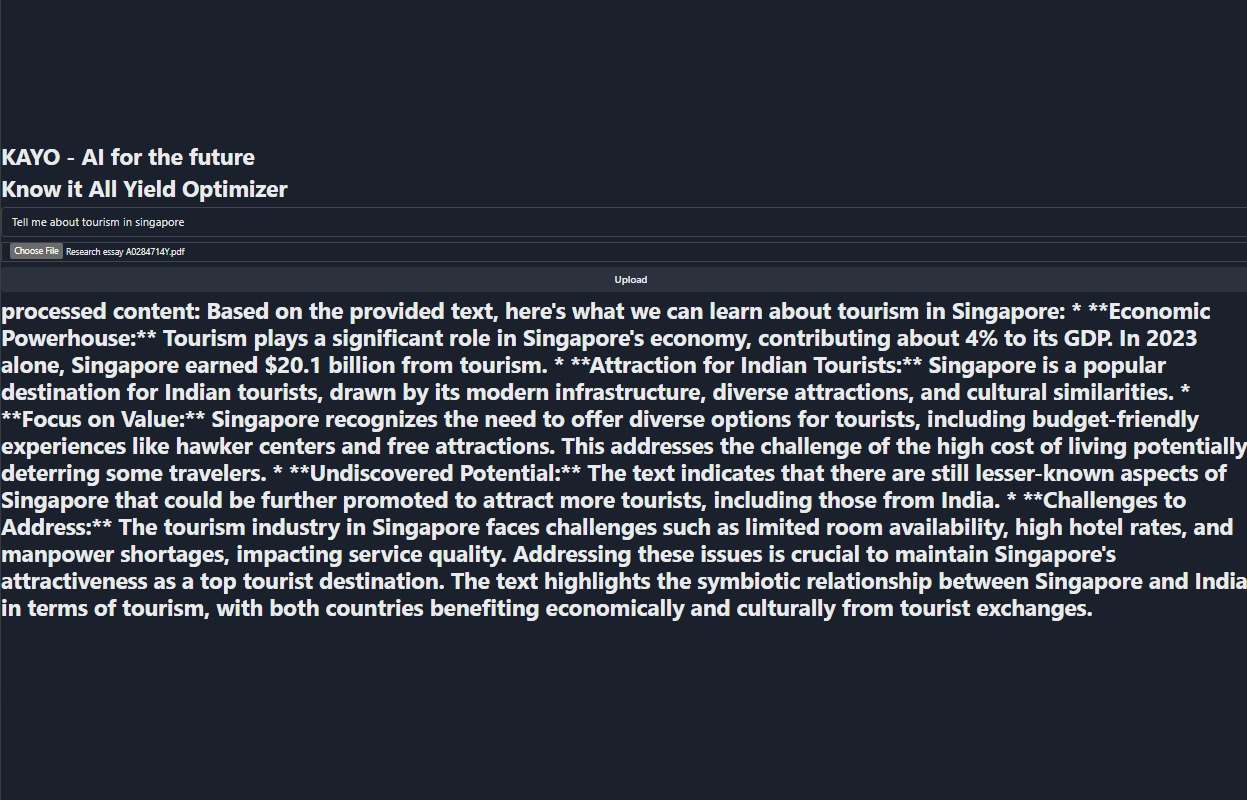
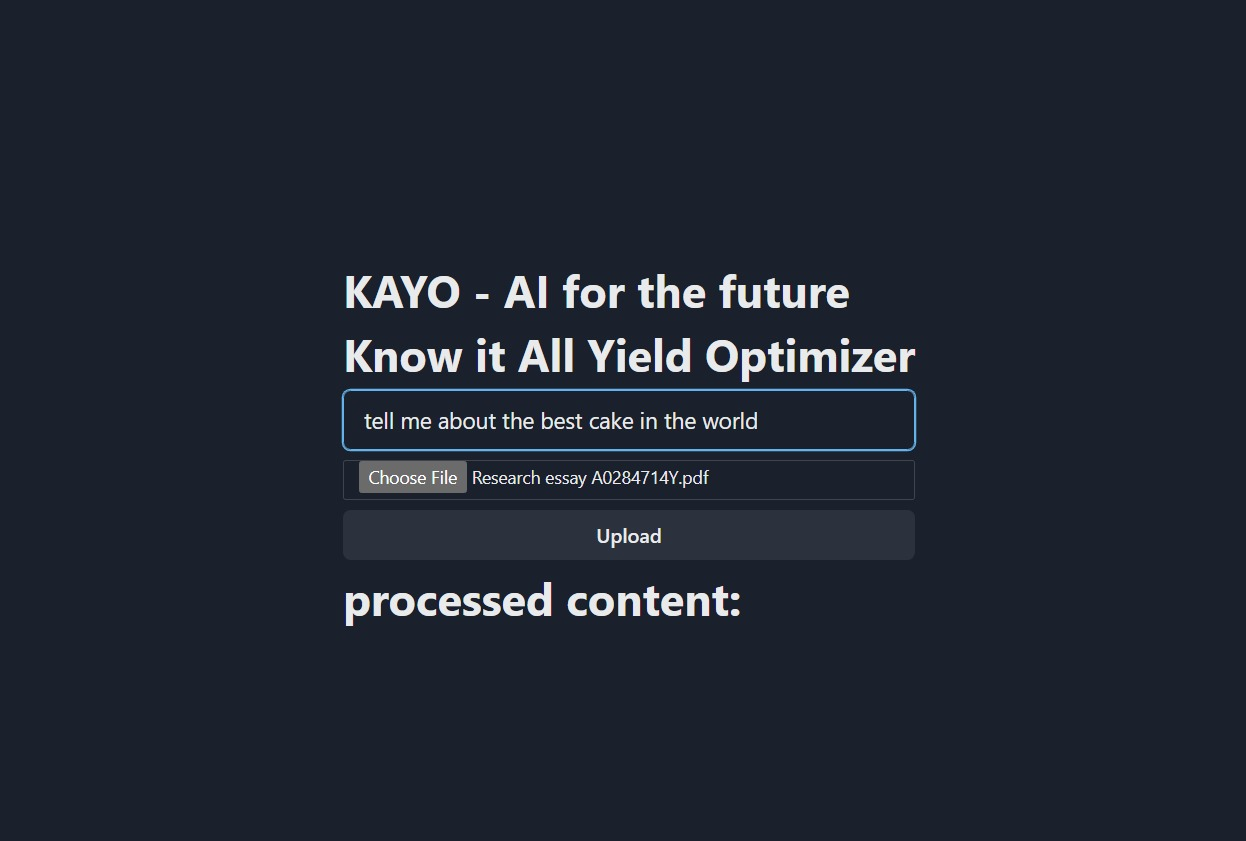
* **Document AI**:
  + KAYO: Designed for usability in academic and research environments with specific enhancements for query accuracy.
  + Document AI: Offers advanced document processing but is more generalized and might not be as finely tuned for academic and research-specific needs.
* **Elastic Enterprise Search**:
  + KAYO: Provides a comprehensive, user-centric solution tailored to academic and research settings.
  + Elastic Enterprise Search: Offers a suite of search tools for various applications but may not be as focused on academic and research-specific use cases.

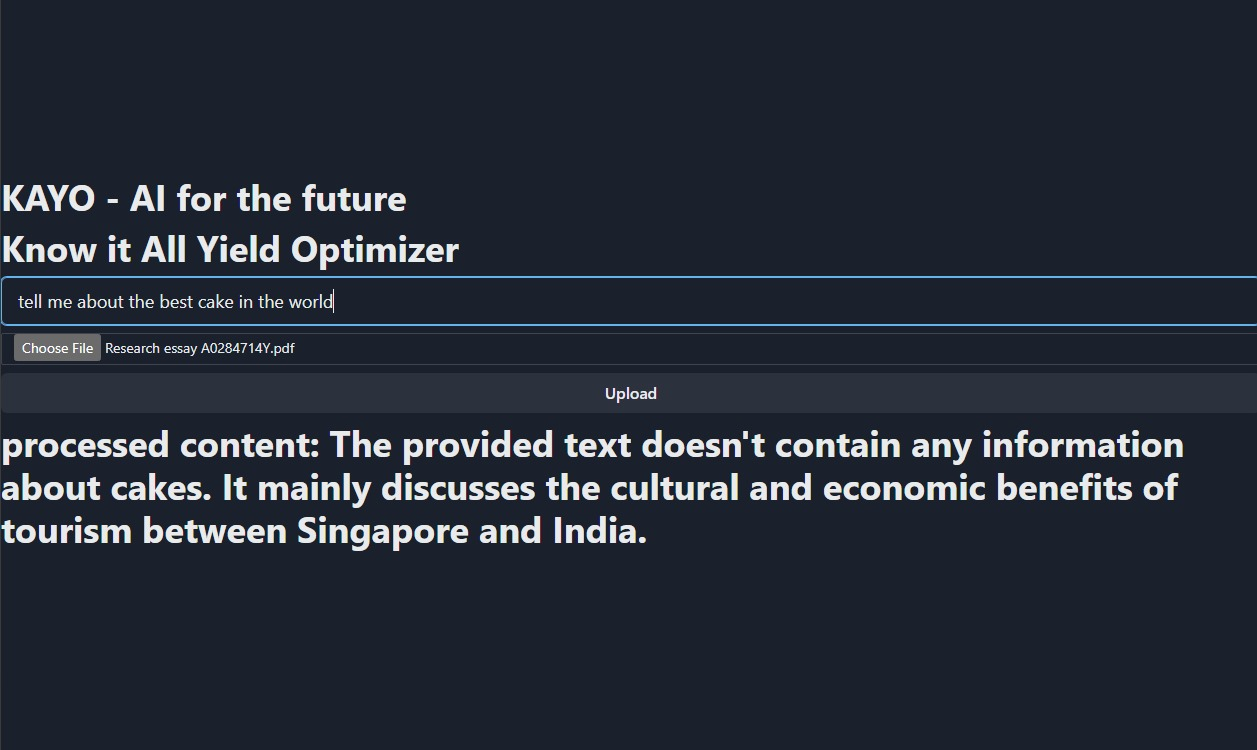
**Development Plan**

| Milestone | Task | Description | Date |
| --- | --- | --- | --- |
| 1 | Research and Learning | 1.Courses  2.Researching and Learning about the tech stack  Researching and learning about AI technologies used  3. Setting up Github  4. Environment Setup | 16 May - 22 May |
| Project Setup and Skeleton | 1.Designing pages and setting up project architecture  Setting up react mockup  2.Chakra and AXIOS setup | 23 May - 27 May |
| LLM and Backend Implementation | 1. Connecting backend with frontend  2. Setting up a document loader and text splitter from Langchain to chunkate the document.  3. Setting up embedding model  4. Setting up similarity search using FAISS  5. Setting up Gemini Generative API | 28 May - 2 Jun |
| Milestone 1 Submission | | | 3 Jun |
| 2 | Improvement on the interface | Improving the UI of the webpage and refining the prompts and output | 4 Jun - 11 Jun |
| Expansion of data types supported | 1.Extending Format support to different types like txt, docs, etc. | 12 Jun - 22 Jun |
| Online Database and Deployment | 1. MongoDB integration  2.Deployment and testing | 23 Jun - 30 Jun |
| Milestone 2 Submission | | | 1 Jul |
| 3 | Setting up custom modes | 1.Research mode  2.Student mode  3. Professional mode | 2 Jul - 10 Jul |
| Feedback Mechanism | Implement a feedback system for users to rate the accuracy and relevance of the responses | 11 Jul - 13 Jul |
| Login Page setup | Authenticate users, store queries | 14 Jul - 28 Jul |
| Milestone 3 Submission | | | 29 Jul |
| REFINEMENT | | | |
|  |  | **FINAL SPLASHDOWN** | 28 Aug |

**Technical Proof of Concept:**

The **technical proof** is as shown below:

1. Webpage:  
   
2. Uploading the document:  
   
3. Entering the prompt:  
   
4. Output:  
   
5. Entering unrelated prompt:  
   
6. Output:



To download and test our product go here - https://drive.google.com/file/d/1RbMspdgQQoknhIHaoOPLuPP6jQMZKMA3/view?usp=sharing