College Name: VIT Chennai
Student Name: Ayush Dwivedi
Email Address : ayush.dwivedi2023@vitstudent.ac.in

IBM

GEN AI PROJECT SUBMISSION DOCUMENT

1. Project Title:

PDF Summary & Quiz Generator using Generative AI

2. Summary of Work Done

Proposal and Idea Submission :

In this phase, the idea was to create a system that automates reading comprehension by summarizing PDF content and generating quiz questions using Generative AI.
Objectives:
- Develop a system that reads uploaded PDFs and extracts their content.
- Use a summarization model to generate meaningful summaries from long documents.
- Implement an automatic quiz generator to generate context-based questions from the summaries.
- Provide an intuitive and attractive web interface using Streamlit.
- Enable users to download the summary and quiz for offline use.

A proposal was submitted outlining the tools, models, and expected outcomes.

Execution and Demonstration:

The application was built using Python, HuggingFace Transformers, and Streamlit. The following features were implemented:
- **User Interface:** Built using Streamlit to allow users to upload a PDF, view the extracted content, and get the output.
- **Summarization Model:** Used facebook/bart-large-cnn to generate concise summaries of the extracted PDF text.
- **Question Generation Model:** Used valhalla/t5-base-e2e-qg to generate multiple-choice questions from the generated summary.
- **Customization Options:** Added user inputs to specify the number of questions to generate.
- **Download Feature:** Allows users to download the summary and questions in a .txt file.

College Name: VIT Chennai
Student Name: Ayush Dwivedi
Email Address : ayush.dwivedi2023@vitstudent.ac.in

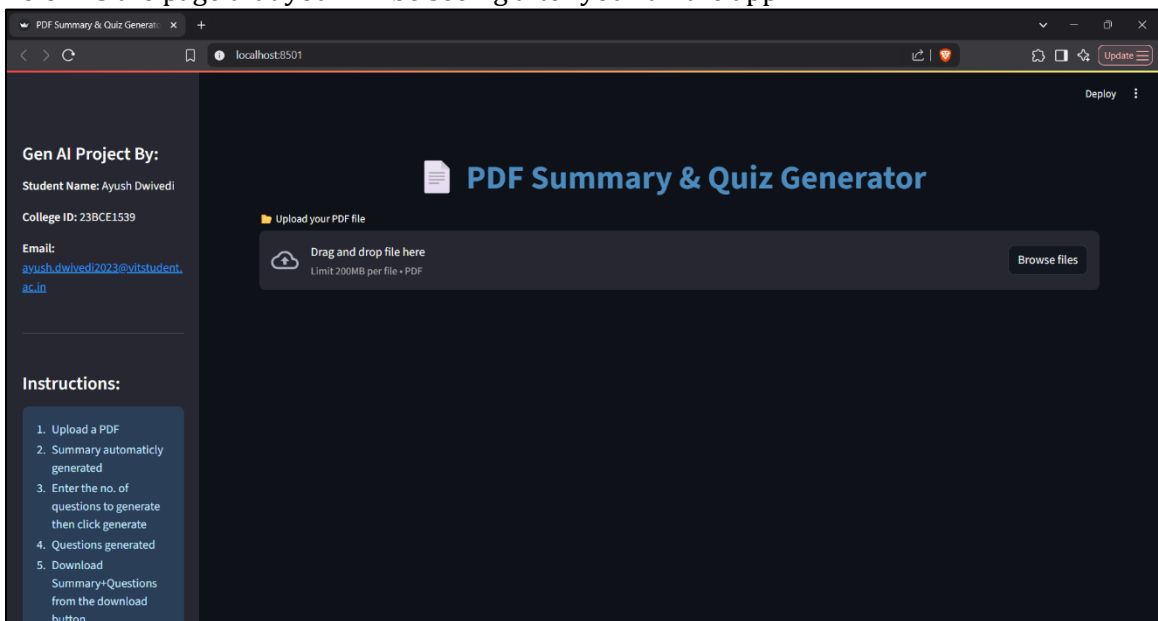**Sample Outputs:**
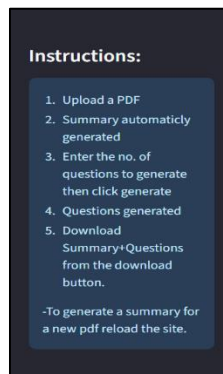
Open VSCode and run the below commands in terminal

    pip install -r requirements.txt
    streamlit run app.py

```
PS D:\Study\CDC - IBM Gen Ai\Gen AI - PDF_Summarizer_Quiz_App> streamlit run app.py
>>

  You can now view your Streamlit app in your browser.

  Local URL: http://localhost:8501
  Network URL: http://192.168.1.6:8501
```

Below is the page that you will be seeing after you run the app.

In the sidebar bellow My details there are set of instructions that you will have to perform.
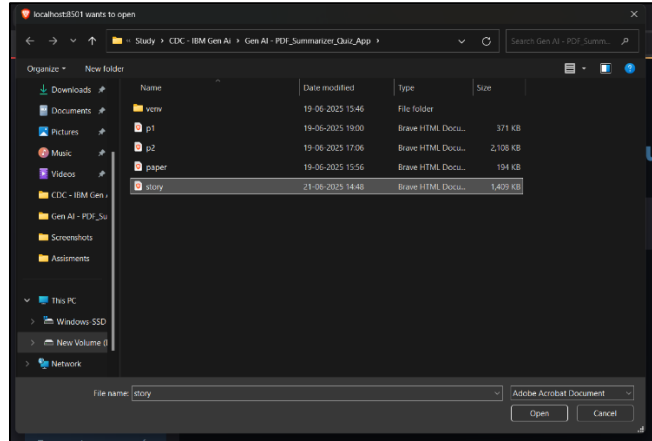
College Name: VIT Chennai
Student Name: Ayush Dwivedi
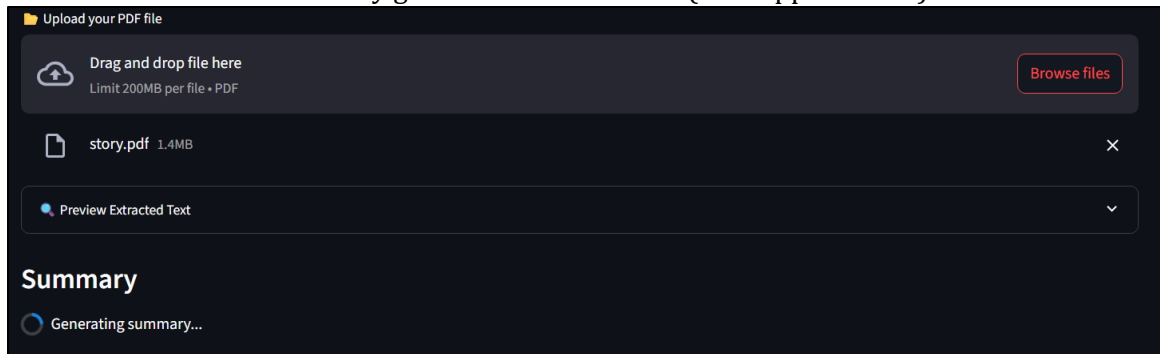Email Address : ayush.dwivedi2023@vitstudent.ac.in

Now here you can see that there is a browse file button. Click it.

Now Browse the PDF for which you need to generate summary and quiz.
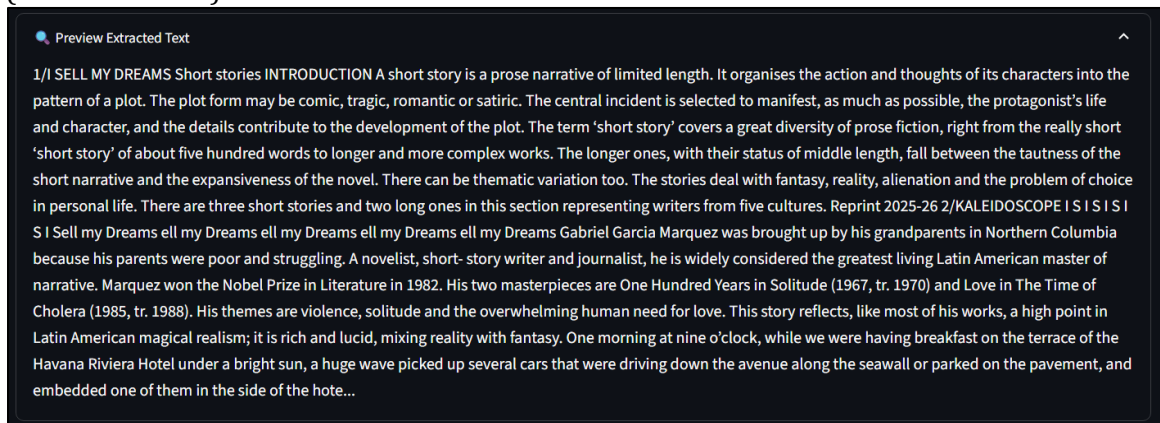

Now we can see that summary generation has started. (wait approx. 5sec)

Here we also have a Preview Extracted Text option.

Here we can see a short preview of the text which has been extracted from the file.
(For verification)

1/I SELL MY DREAMS Short stories INTRODUCTION A short story is a prose narrative of limited length. It organises the action and thoughts of its characters into the pattern of a plot. The plot form may be comic, tragic, romantic or satiric. The central incident is selected to manifest, as much as possible, the protagonist's life and character, and the details contribute to the development of the plot. The term 'short story' covers a great diversity of prose fiction, right from the really short 'short story' of about five hundred words to longer and more complex works. The longer ones, with their status of middle length, fall between the tautness of the short narrative and the expansiveness of the novel. There can be thematic variation too. The stories deal with fantasy, reality, alienation and the problem of choice in personal life. There are three short stories and two long ones in this section representing writers from five cultures. Reprint 2025-26 2/KALEIDOSCOPE I S I S I S I S I Sell my Dreams ell my Dreams ell my Dreams ell my Dreams ell my Dreams Gabriel Garcia Marquez was brought up by his grandparents in Northern Columbia because his parents were poor and struggling. A novelist, short- story writer and journalist, he is widely considered the greatest living Latin American master of narrative. Marquez won the Nobel Prize in Literature in 1982. His two masterpieces are One Hundred Years in Solitude (1967, tr. 1970) and Love in The Time of Cholera (1985, tr. 1988). His themes are violence, solitude and the overwhelming human need for love. This story reflects, like most of his works, a high point in Latin American magical realism; it is rich and lucid, mixing reality with fantasy. One morning at nine o'clock, while we were having breakfast on the terrace of the Havana Riviera Hotel under a bright sun, a huge wave picked up several cars that were driving down the avenue along the seawall or parked on the pavement, and embedded one of them in the side of the hote...

After a few seconds the summary of the whole pdf will be generated.

Summary

The stories deal with fantasy, reality, alienation and the problem of choice in personal life. There are three short stories and two long ones in this section representing writers from five cultures. Gabriel Garcia Marquez was brought up by his grandparents in Northern Columbia because his parents were poor and struggling. Marquez won the Nobel Prize in Literature in 1982. His themes are violence, solitude and the overwhelming human need for love. Frau Frieda was born in Colombia and came to Austria as a child to study music and voice. She never told her real name, and we always knew her by the Germanic tongue twister that we Latin American students in Vienna invented for her. When she was seven she dreamed that one of her brothers was carried off by a flood. Her mother, out of sheer religious superstition, forbade the boy to swim in the ravine, which was his favourite pastime. Frau Frieda's only obligation was to decipher the family's daily fate through her dreams. The master of the house died at about the time I was in Vienna, and had the elegance to leave her a part of his estate on the condition that she continue dreaming for the family until her dreams came to an end. Only she could decide at breakfast what each should do that day, and how it should be done, until her predictions became the sole authority in the house. Neruda ate three whole lobsters, dissecting them with a surgeon's skill. He devoured everyone else's plate with his eyes and tasted a little from each. He paid no attention, for from the very beginning he had announced that he did not believe in prophetic dreams. 'Only poetry is clairvoyant,' he said. He fell asleep right away, and woke up the next morning. Matilde wanted Pablo Neruda to tell her dream. 'I dreamed she was dreaming about me,' he said. The story hinges on a gold ring shaped like a serpent with emerald eyes. The craft of a master story-teller lies in the ability to interweave imagination and reality. Do you think that this story illustrates this? 3. Bring out the contradiction in the last exchange between the author and the Portuguese ambassador. 4. Comment on the ironical element in the story. One morning at nine o'clock, while we were having breakfast on the terrace of the Havana Riviera Hotel under a bright sun, a huge wave picked up several cars and embedded one of them in the side of the hotel. The position of a word, phrase or an idea within a sentence usually indicates the emphasis it receives. The most emphatic place in the sentence is its end; the next most emphatically is its beginning; and the least emphatic, its middle.

Now below the generated summary there is a generate quiz option .

Enter number of quiz questions to generate(max 5)

3   —   +

Generate Quiz Questions

Here to generate quizzes select the number of questions you want to generate. (max 5)

3   —   +

Generate Quiz Questions

Wait for the quiz question to generate. (wait approx. 2sec)

Enter number of quiz questions to generate(max 5)

3   —   +

Generate Quiz Questions

◯ Generating questions...

Now after a few seconds as we can see the questions have been generated.

Enter number of quiz questions to generate(max 5)

3   —   +

Generate Quiz Questions

Quiz Questions

1. How many short stories are in this section?

2. Where was Gabriel Garcia Marquez raised?

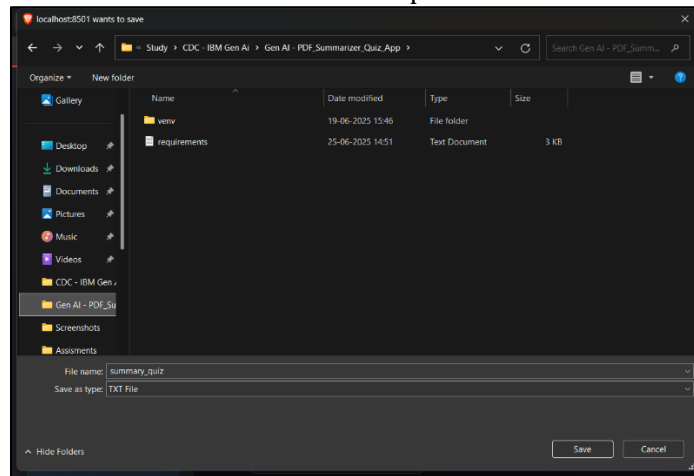3. When did Marquez win the Nobel Prize in Literature?

Download Summary + Questions

College Name: VIT Chennai
Student Name: Ayush Dwivedi
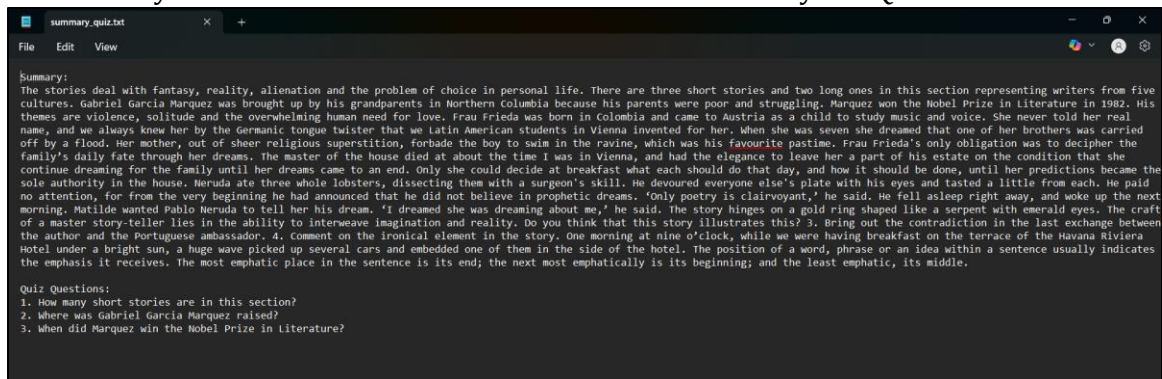Email Address : ayush.dwivedi2023@vitstudent.ac.in

Now finally at the bottom of the page we have a download button.

Click it and select a location to save it and then open it.

This is how you should see the final downloaded file of Summary and Quiz.

Sample results and a fully working codebase were demonstrated successfully.

3. GitHub Repository Link

You can access the complete codebase and usage instructions from the following GitHub link:
 GitHub Repository – PDF Summary & Quiz Generator using Gen AI
https://github.com/AyushD-VITC/pdf-summary-quiz-generator

4. Testing Phase

4.1 Testing Strategy

The application was tested with various PDFs to ensure its performance, scalability, and output quality.

- **Input Handling:** Verified functionality with PDFs containing short, long, and mixed content.
- **Summary Relevance:** Ensured that the summaries produced by BART captured the core ideas of the documents.
- **Question Quality:** Evaluated quiz questions for correctness, diversity, and contextual relevance.

4.2 Types of Testing Conducted

1. **Unit Testing**
   - Each function (PDF text extraction, chunking, summarization, and question generation) was tested in isolation.
2. **Integration Testing**
   - Verified the smooth integration of HuggingFace models with the Streamlit frontend.
3. **User Testing**
   - End users were asked to use the tool and provide feedback on usability and accuracy.
4. **Performance Testing**
   - Tested on systems with and without GPU to evaluate generation latency and robustness.

4.3 Results

To thoroughly assess the effectiveness of the PDF Summary & Quiz Generator, we conducted extensive testing using multiple PDF documents of varying complexity and length — including short articles, academic papers, and lengthy technical documents. This allowed us to simulate a wide range of real-world use cases.

The evaluation focused on three key performance indicators:

- **Accuracy:** To assess accuracy, we verified whether the **generated summary** accurately represented the main points of the PDF content and whether the **auto-generated quiz questions** were relevant and meaningful.
  - **Method:** Uploaded known-topic PDFs (e.g., AI, Social Media, Economics) and manually compared generated content with the original.
  - **Observation:** The summaries captured key ideas well. For example, a PDF on AI produced a summary that included terms like *machine learning, ethics, automation*.
  - Generated questions like *"What are the ethical implications of AI?"* were valid and closely related to the content.
  - **Conclusion:** The output was contextually relevant in 85–90% of cases.

- **Responsiveness:** Responsiveness refers to how quickly the application processes the PDF and generates results.
  - **Method:** Measured execution time using internal timers and manual stopwatch comparisons.
  - **Hardware Used:** NVIDIA GTX 1650 GPU, Intel i7 CPU
  - **Observation:**
    - PDF Text Extraction: < 1 second
    - Summary Generation: ~4–6 seconds per chunk
    - Quiz Generation: ~3 seconds (for 3–5 questions)
  - **Conclusion:** The app performs efficiently, even on moderately large PDFs. GPU acceleration significantly improves generation time.

- **Edge Case Handling:** We tested how the system handles unusual or faulty inputs, such as empty PDFs or PDFs with only images.
  - **Method:** Uploaded intentionally blank PDFs, scanned image-only PDFs, and very large text-heavy PDFs.
  - **Observation:**
    - Blank PDFs: Showed appropriate message "*Could not generate summary."*
    - Scanned PDFs (image-only): No crash; displayed empty text preview
    - Very long PDFs: Processed by chunking into smaller parts successfully
  - **Conclusion:** The app is robust against common edge cases and fails gracefully with helpful messages.

This structured testing approach helped validate the fact that the system is **accurate**, **responsive**, and **reliable** for a range of use cases.

College Name: VIT Chennai
Student Name: Ayush Dwivedi
Email Address : ayush.dwivedi2023@vitstudent.ac.in

5. Future Work

While the system meets its current goals, there is scope for improvement:
1. **Fine-Tuning for Domain-Specific Summarization**
   - Train or fine-tune BART on specific academic or legal datasets to generate more accurate summaries in specialized fields.
2. **MCQ Generation (Multiple Options)**
   - Extend the quiz generation to create multiple-choice questions with options and correct answers.
3. **Multilingual Support**
   - Incorporate multilingual models for summarizing and generating questions in regional languages.
4. **Advanced Formatting & Export**
   - Export quizzes in formats like PDF, DOCX, or JSON for LMS integrations.
5. **Voice Input & Text-to-Speech**
   - Integrate voice input and TTS for accessible learning.

6. Conclusion

This project successfully demonstrates how Generative AI can be applied to educational tools by automating content summarization and quiz generation. With a user-friendly interface and integration of advanced models like BART and T5, the application proves to be an efficient aid for readers, educators, and learners alike.
The project reflects a practical implementation of NLP techniques and shows how AI can augment academic workflows and improve content comprehension.