

Building a GenAI-Powered Document Summarization Tool

Objective:

Develop a basic GenAI application that accepts user-uploaded documents (PDF or DOCX), summarizes the content using a language model, and presents the summary in a user-friendly format.

Scope of Work:

1. Document Ingestion:

- Implement file upload functionality.
- Extract text from PDF and DOCX formats.

2. Text Preprocessing:

- Clean and tokenize the extracted text.
- Handle formatting issues, special characters, and large text blocks.

3. GenAI Integration:

- Use a GenAI model (e.g., OpenAI GPT, Azure OpenAI, or Hugging Face models) to generate a concise summary.
- Ensure the summary maintains the original document's intent and tone.

4. User Interface:

- Build a simple web interface using Flask or Streamlit.
- Allow users to upload documents and view/download the summary.

5. Bonus Tasks (Optional):

- Add multi-language support.
- Include a feature to generate bullet-point summaries or executive briefs.
- Implement basic analytics (e.g., word count, sentiment analysis).

Deliverables:

- Source code (GitHub repository or ZIP)
- A short demo video or screenshots
- Documentation covering setup, usage, and architecture

Deadline: [August 20, 2025]