A. Overall Approach

The primary goal of this project is to create a chatbot capable of answering questions about a company's products and information. The approach involves the following steps:

1. Extract Text from PDF: The text content is extracted from a given PDF document using the PyPDF2 library.
2. Process Text: The extracted text is processed to identify prompts (questions) and corresponding responses.
3. Setup Question-Answering Model: A pre-trained BERT model (`distilbert-base-uncased-distilled-squad`) is used for the question-answering task.
4. Build Chatbot Interface: Streamlit is used to create a simple web interface where users can input questions and receive answers from the chatbot.

B. Frameworks/Libraries/Tools Used

1. Streamlit: Used to build the web application interface.
2. Transformers (Hugging Face): Provides the pre-trained BERT model and tokenizer for the question-answering task.
3. PyPDF2: Used to extract text from PDF files.
4. Torch: Used to check GPU availability and support model execution.

C. Problems Faced and Solutions

1. Text Extraction Issues:
   Extracting clean text from PDFs can be challenging due to the varying structure of PDFs.
   This was mitigated by replacing multiple newlines and special characters with spaces.

   - Solution: Added a text cleaning step to ensure continuous and readable text.

2. Handling Large Contexts:
   BERT models have a limit on the context size they can handle.

   - Solution: Limited the context to a reasonable size and used the most relevant parts of the text.

3. Performance:
   Ensuring the chatbot responds quickly.

   - Solution: Utilized GPU if available to speed up model inference.

D. Future Scope

1. Enhanced Text Processing:
   Improve the text extraction and processing to handle more complex PDF structures.

2. Contextual Understanding:
   Implement methods to maintain conversation context across multiple user queries.

3. Additional Features**:
   - Add functionality for extracting and answering questions based on tables and figures in the PDF.
   - Implement multi-language support for a broader user base.

4. Scalability:
   Deploy the chatbot using scalable cloud services to handle more users simultaneously.

5. User Interface Improvements:
   Enhance the UI to provide a more interactive and user-friendly experience.

By addressing these areas, the chatbot can become more robust, accurate, and user-friendly, providing better support for users seeking information about a company's products and services.