**Challenges :**

1. **Reliable and Working LLM API**:
   * **Challenge**: Anyscale updated its policies, limiting access to its free LLM API. Hugging Face, a popular LLM API, has most of its models unavailable for Inference endpoints.
   * **Solution**: Utilized the Eden AI platform, which provides LLM API with some free credits, overcoming the limitations posed by other platforms.
2. **Text Processing**:
   * **Challenge**: Extracting and preprocessing text from PDF documents can be complex, especially when dealing with different layouts, fonts, and encoding issues.
   * **Solution**: Used PyPDF2 for text extraction and implemented text preprocessing steps such as removing extra whitespace and handling line breaks to ensure clean and usable text for query processing.

**Scalability Consideration :**

1. **Support for Additional Document Formats**:
   * Extending support to other document formats like DOCX, TXT, and HTML will make the application versatile for a broader audience.
2. **User Authentication and Personalization**:
   * Implementing user authentication can provide personalized experiences and save user history, enhancing user engagement.
3. **Caching Mechanism**:
   * Implementing a caching mechanism to store and reuse responses for repeated or similar
4. **Integration with Additional APIs**:
   * Incorporating other language model APIs will offer a variety of response options, ensuring the application remains robust and up-to-date with the latest NLP advancements.

**Potential Future Improvements**

1. **Interactive pdf navigation**, where users can click on highlighted content and jump directly to corresponding section.
2. **OCR integration**, so that users can chat with some handwritten content directly
3. **Caching Mechanism**: Implement caching for faster responses.
4. **Mobile Compatibility:** Ensure full responsiveness for mobile use.