PROBLEM STATEMENT 2
Legal Document Summarizer

Problem Statement

Develop a Legal Document Summarizer tailored for NGOs and citizens, capable of simplifying complex legal policies — such as environmental laws and human rights acts — into clear, concise, and accessible summaries. The system should accurately extract key points while preserving the legal context, making crucial information easier to understand for non-experts. Additional features like multilingual support, keyword highlighting, and actionable insights would enhance usability and impact.

Type of Input

- Government legal documents, NGO policy papers in various formats (pdf, word etc)

Expected Output

- A tool which accurately summarises various document formats (e.g., PDF, Word, text) and extracts the most relevant information, avoiding irrelevant details or biased interpretations.
- The key information must be displayed in a format easy to understand and read
- Integration of Technologies like BERT (for summarization), TextBlob (for readability), NLP Question-Answering Model

Must haves

- **Accurate Summarization:**Extract key points while maintaining legal context and avoiding misinterpretation.
- **Context Preservation:**Ensure that legal nuances, clauses, and responsibilities aren't lost in the summarization process.
- **Keyword Highlighting:**Automatically detect and highlight crucial legal terms, dates, names, and obligations.
- **Input Flexibility:**Accept multiple formats (PDF, DOCX, scanned documents, or plain text).
- **Low Latency:**Process documents quickly to generate summaries in near real-time.

Should Haves from the Prototype

- **Multilingual Support:**Translate summaries into regional languages for broader accessibility.
- **User-Friendly Interface:**Clean and intuitive UI for uploading documents and viewing summaries.

- **Output Options:** Provide summaries in structured formats like bullet points, FAQs, or action items.
- **Source Referencing:** Include citations or references to original clauses for fact-checking.

Additional Considerations

● Scalability: Solutions should be designed to handle increasing volumes of data and adapt to new input types.

● Data Quality: Emphasis on preprocessing, data cleaning, and augmentation techniques to improve model performance.

● Ethical & Privacy Issues: Ensure all prototypes comply with data privacy regulations and ethical guidelines.

● Documentation & Reproducibility: Clear code documentation, experiment logs, and version control for reproducibility.