# POC Project Documentation

---

**1. Project Title**

Automated Legal Document Summarizer

---

**2. Team Member Details**

| Name | Email | Role |
|---|---|---|
| Nandan Javagal | nandanjavagal444@gmail.com | Intern |

---

**3. Problem Statement**

Legal professionals and researchers face challenges while processing lengthy legal documents. Manually reading and extracting key legal facts, judgment reasoning, and statutory references is time-consuming, inefficient, and error-prone.

---

**4. Objective**

To develop an AI-powered web application that takes legal court documents (PDFs) and automatically extracts key legal facts, reasoning, and statutory references, summarizing them in clear legal English.

---

**5. Proposed Solution**

- Extract PDF text using pdfplumber
- Clean and pre-process the legal text
- Split large text into sentence-based chunks using NLTK
- Use FLAN-T5 (google/flan-t5-large) for enhanced prompt-based summarization
- Design legal assistant-style prompts for meaningful outputs
- Build UI in Streamlit to upload, preview, summarize, and download
- Support multiple levels of summary detail (Concise, Detailed, Comprehensive)

---

**6. Tech Stack**

- Language: Python
- Libraries: Streamlit, transformers, pdfplumber, nltk
- Model: google/flan-t5-large
- Other Tools: Hugging Face, PyTorch, Sentence Tokenizer
- Deployment: Streamlit Web App

**7. Dataset Details**

- No fixed dataset required; user-uploaded legal PDFs (court judgments, writs)
- Extractive input using pdfplumber
- Evaluation based on actual use cases

---

**9. Model/Logic Explanation**

- Prompts simulate legal assistant tasks (extract reasoning, facts, citations)
- Summarization done via beam search, top-p sampling, and length penalties
- Post-processing to clean up noisy output and ensure clarity

---

**10. Input & Output Format**

- **Input**: PDF legal documents
- **Output**: Clean legal summaries including:
    - Key Legal Facts
    - Judgment Reasoning
    - Statutory References
    - Output can be viewed in the UI and downloaded as a .txt file