

Abstract: Legal Document Summarization System

Legal documents are often long and complex, making it difficult for lawyers, businesses, and individuals to quickly understand key information. Our project aims to automate the process of summarizing legal documents using AI-powered OCR (Optical Character Recognition) and a custom language model.

The system works in three main steps:

1. **Text Extraction:** We use Tesseract OCR to convert handwritten or printed legal documents into digital text.
2. **Text Processing & Cleaning:** The extracted text is cleaned and prepared for summarization.
3. **Summarization Model:** A custom-trained AI model generates concise and legally accurate summaries using Retrieval-Augmented Generation (RAG) to improve accuracy.

END USERS

- **Lawyers & Legal Professionals** – Saves time by summarizing long legal documents, helping them focus on important details.
- **Companies & Business Owners** – Quickly understand contracts, agreements, and legal risks without needing a lawyer.
- **Law Students & Researchers** – Helps in studying legal cases by providing short, clear summaries.

KEY PROBLEMS

- Time-Consuming Legal Reviews
- Complex Legal Language
- High Cost of Legal Consultation

Websites Similar to Our Legal Document Summarization Project

- [Legalyze.ai](#)
- [LegalZoom's Doc Assist](#)
- [Case Digest](#)

