

AI Powered Legal Document Analyser

Prof. Tejashree pangare, Tanvi Patil, Pratham Surve

Computer Engineering Pillai HOC College of
Engineering and Technology Rasayani, India

Abstract- The AI-Powered Legal Document Analyser is an Android application created to assist users in understanding complicated legal documents including contracts, agreements, and legal notices. Many legal documents contain complex terminology that is hard for non-expert professionals to interpret. This application enables users to upload or scan documents, after which Optical Character Recognition (OCR) retrieves the text and Natural Language Processing (NLP) techniques examine the content to detect key clauses, responsibilities, potential risks, and time limits. The application also includes a legal chatbot that responds to user queries, generates easy summaries of the document. Built using Java and XML for the Android interface and Firebase for safe data storage, the system maintains user privacy while offering fast and dependable legal insights. This project aims to make legal information easier to access, affordable, and easy to understand for individuals, students, and small businesses.

Keywords: Artificial Intelligence, Legal Document Analysis, Natural Language Processing (NLP), Generative AI, Optical Character Recognition (OCR), Android Application, Legal Chat- bot, Firebase, Contract Summarization, Legal Tec.

I. INTRODUCTION

In the modern digital era, legal documents such as contracts, agreements, and notices are becoming lengthy and complicated. Many people find it difficult to understand important details like penalties, obligations, and deadlines because legal language is often technical and complicated. Consulting a lawyer for every document can be costly and time-consuming. With recent developments in Artificial Intelligence (AI), technologies such as Natural Language Processing (NLP) and Optical Character Recognition (OCR) can now read, analyse, and summarize difficult text.

The AI-Powered Legal Document Analyser uses these technologies to help users upload or scan legal documents and easily understand important information. The system is developed as an Android application with secure Firebase database integration, the system design to make legal documents simpler, more accessible, and simpler to interpret for everyone. This application contains a legal chatbot that allows users to ask questions about the document and receive clear, easy-to-understand answers.

Background

In today's digital era, legal documents such as contracts, agreements, and notices have become longer and more complicated. The technical language used in these documents makes it difficult for general users to understand critical clauses related to penalties, obligations, rights, and deadlines. Therefore consequently, many individuals and small businesses experience confusion or potential legal risks due to misunderstanding important information. Professional legal advice although reliable, is often expensive and time consuming. The AI- Powered Legal Document Analyser is an Android application created to simplify the understanding of such legal documents.

Using Optical Character Recognition (OCR) technology to retrieve text from scanned and uploaded documents such as PDFs and images. It then uses Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques to examine the content, detect key clauses, indicate potential risks, and produce simple summaries. Additionally, the application includes a integrated legal chatbot that enables users to ask questions about the document and receive clear, human readable responses. Developed using Java/XML for the front-end and Firebase for secure user authentication and data storage, the system ensures privacy, reliability,

and smooth performance. In conclusion, this project aims to make legal information more accessible, affordable, and understandable for everyone, increasing dependency on legal experts for simple document understanding.

II. MOTIVATION

The primary goal of creating the AI-Powered Legal Document Analyser is to enhance the comprehension of legal information and make it available to all. A significant number of individuals sign legal documents without fully grasping their provisions, which can lead to misunderstandings or legal issues down the line. This initiative aims to bridge that gap by leveraging Artificial Intelligence to clarify intricate legal language.

a] Lack of Legal Awareness

Several individuals do not have legal awareness and find it difficult to understand complex terms used in contracts and agreements. Important details such as penalties, renewal conditions, and obligations are frequently ignored. This system helps users clearly identify and understand such important clauses.

b] High Cost of Legal Consultation

Hiring a lawyer for examining every document can be costly or time consuming. Small businesses, students, and freelancers may always be able to afford expert legal advice. This application provides quick initial insights, helping users reduce time and cost.

c] Use of AI for Social Benefit

With developments in OCR, AI and NLP, technology can now analyse and interpret text efficiently. This project aims to responsibly use these technologies to enhance transparency, encourage legal awareness, and support better decision making without substituting legal professionals.

preparing the ,research can train advanced AI models on them, researchers can developed systems that give more accurate and meaningful answers based on the content of the document.

S. Patel and D. Nair[1]: S. Patel and D. Nair developed a system where documents are first cleaned and processed using NLP techniques. After that, a Generative AI model like GPT or T5 is trained to read the document and answer questions based on its content. The system gave accurate answers, but it needed a lot of data and high computing power to work properly.

K. Tanaka and H. Suzuki[2]: They used NLP models such as BERT and T5 to process legal documents and understand their content. They applied transformer-based techniques to generate clear and meaningful summaries of lengthy legal texts. In addition, they fine-tuned a generative AI model so that it could accurately answer user questions based on the document's content.

Y. Choudhary, A. Bhatt, and N. Saxena[3]: They used NLP techniques to clean text data and extracted important features using TF-IDF. They trained a Random Forest model on labelled data to classify and detect patterns. Similarly, in an AI-Powered Legal Document Analyser, this approach can be used to classify legal clauses and identify important or risky terms in documents automatically.

R. Zhao, L. Chen, and Y. Wang[4]: They developed a system where scanned legal documents are first converted into editable text using OCR technology. After extracting the text, NLP techniques such as breaking the text into words (tokenization), identifying important names and terms (Named Entity Recognition), and classifying the content are applied to analyse the document and generate useful insights.

Sharma and P. Kulkarni[5]: They proposed a system where contracts are first processed using NLP techniques such as tokenization and named entity recognition to identify important terms, parties, and legal entities. After preprocessing, an AI chatbot powered by transformer-based models analyses the

III. LITERATURE REVIEW

Recent research indicates that when Natural Language Processing (NLP) is integrated with Generative AI models, computers can understand documents more better. By preprocessing and

clauses and answers user questions based on the contract content.

L. Chalkidis and I. Androutsopoulos[6]: They explored the use of NLP techniques for processing legal documents and applied transformer-based models to generate short and meaningful summaries. Their approach showed that advanced language models can understand complex legal text and produce clear, concise summaries while preserving important information.

I. Chalkidis, M. Fergadiotis, and I. Androutsopoulos[7]: They studied the use of transformer-based models such as BERT and T5 for processing legal documents. In their approach, legal texts were first analysed using NLP techniques, and then advanced transformer models were used to generate clear and meaningful summaries.

IV. LIMITATIONS

Although the AI-Powered Legal Document Analyser helps users understand complex legal documents, it has certain limitations. Its performance depends on OCR accuracy, quality of training data, and the ability of AI models to interpret legal language correctly. The system provides assistance but cannot fully replace professional legal advice.

1. Dependence on OCR Accuracy:

The system relies on OCR to extract text from scanned or image-based documents. If the document quality is poor, blurred, or handwritten, the extracted text may contain errors. These mistakes can affect the accuracy of analysis and summaries.

2. Limited Understanding of Complex Legal Context:

Although AI models like NLP and transformer-based systems can analyse text, they may not fully understand highly complex or rare legal scenarios. Some clauses may require deep legal interpretation that only experienced legal professionals can provide

3. Dataset Dependency:

The performance of the system depends on the quality and size of the training data. If the AI model is not trained on diverse legal documents from different domains, it may not perform equally well for all types of contracts or jurisdictions.

4. High Computational Requirements:

Transformer-based models such as BERT or T5 require significant computing power and memory. This can make deployment on mobile devices challenging and may require cloud support, increasing operational costs.

5. Not a Replacement for Legal Professionals:

The system provides guidance and summaries but cannot replace professional legal advice. It may miss subtle legal risks or provide generalized interpretations that should still be verified by a qualified lawyer for critical decisions.

V. EXISTING SYSTEM

Many existing systems only provide basic features like storing legal documents or searching for keywords. Traditional Document Management Systems (DMS) allow users to upload and manage files, but they do not analyse or summarize the content. Similarly, OCR tools can convert scanned documents into text, but they cannot understand the meaning of the legal language or answer user questions. Some advanced AI platforms like ROSS Intelligence and LawGeex help in contract review and legal research. However, these systems are mostly commercial, expensive, and designed for large law firms. They are not easily accessible for students, individuals, or small businesses. Certain research projects have used machine learning to classify legal clauses or detect risks, but these systems are usually limited to specific datasets or legal areas. Most of them also do not provide an interactive chatbot for real-time user queries.

Compared to these systems, the AI-Powered Legal Document Analyser offers a mobile-based, affordable, and user-friendly solution. It combines OCR, NLP-based summarization, and a chatbot

interface to help users quickly understand legal documents in a simple way.

VI. PROBLEM STATEMENT

Legal documents are usually long and written in complicated language, which makes them difficult for common users to understand. People often struggle to identify important details such as penalties, responsibilities, and deadlines. Students and small organizations especially find it hard to interpret such documents correctly. Most existing digital tools only allow users to store documents or search for keywords. They do not explain clauses clearly or provide interactive assistance. Although hiring a lawyer gives accurate guidance, it can be expensive and time-consuming. Many AI-based tools available today are limited to specific domains, languages, or are too costly for regular users. Therefore, there is a need for an AI-powered system that can automatically extract text using OCR, analyse it using NLP and Generative AI, and provide simple summaries and explanations. The system should also allow users to ask questions in normal language and receive clear answers, while keeping their data secure through a reliable backend system like Firebase.

VII. PURPOSED ARCHITECTURE

The architecture of the AI-Powered Legal Document Analyser starts with the user uploading or scanning a document through the Android application. The OCR engine extracts the text from the document, which is then processed by the AI engine using NLP and Generative AI to detect important clauses, risks, deadlines, and generate summaries. The analysed results are displayed on the app through a results screen and chatbot interface, while Firebase ensures secure user authentication and safe storage of documents and data.

VIII. METHODOLOGY

The working of the AI-Powered Legal Document Analyser begins with user authentication through Firebase, which ensures that each user has a secure and personalized account. After logging in, the user

can upload a legal document in PDF or image format, or scan it directly using the mobile camera. The uploaded file is securely stored in Firebase Cloud Storage, and its basic details are saved in the database. Once the document is stored, the system uses an OCR service such as Google Cloud Vision to extract readable text from the file. The extracted text is cleaned and organized before being sent to the AI engine. The AI engine, powered by Natural Language Processing (NLP) and Generative AI models, analyses the text to identify important clauses, obligations, penalties, deadlines, and other critical information. It also generates short and easy-to-understand summaries to help users quickly grasp the document's meaning.

The system then classifies detected clauses into categories such as termination, payment, confidentiality, liability, and jurisdiction. A built-in legal chatbot allows users to ask questions about the document, and it provides clear, natural-language responses based on the analysed content. All extracted data, summaries, and chatbot responses are securely stored in Firebase under the respective user ID for privacy and easy retrieval. Finally the Android application displays the summarized content, highlighted clauses, and chatbot responses in an interactive format. The system is tested for accuracy, speed, and reliability to ensure smooth performance. Future improvements may include multilingual support, enhanced visualization, and more advanced AI features.

IX. RESULT ANALYSIS

The AI-Powered Legal Document Analyser was tested using different legal documents such as contracts and agreements to check its performance and accuracy. The OCR feature was able to convert most scanned and PDF documents into readable text, although a few small mistakes occurred when the image quality was poor. The NLP and AI models successfully detected important parts of the document, including clauses related to termination, payment conditions, and important deadlines. The system also produced short and clear summaries that made long legal documents easier to understand for users. The chatbot responded to user

questions and explained legal terms in simple language. Firebase was used to securely store documents and allowed quick retrieval whenever needed. The application interface was simple and easy to use, so users could upload and analyse documents without difficulty. Overall, the project proved that using AI can make the process of understanding legal documents faster, easier, and more efficient. In addition, the system helped reduce the time required to manually read and analyse long legal documents. Users were able to quickly find important information without going through the entire document. The application also improved accessibility for people who are not familiar with complex legal language.

- 1] Home page
- 2] Features
- 3] Context Conversation

X. CONCLUSION AND FUTURE SCOPE

The AI-Powered Legal Document Analyser helps simplify the process of reviewing legal documents by using technologies like Natural Language Processing (NLP) and Machine Learning (ML). The system automatically extracts important clauses, detects risks, checks compliance, and generates summaries, reducing manual effort and saving time. This makes legal document analysis faster, more accurate, and more efficient. Overall, the project shows how AI can improve traditional legal processes by making them more digital, secure, and user-friendly. With future improvements and advanced AI integration, the system has strong potential to become an effective tool in the growing field of Legal Technology.

In the future, the AI- Powered Legal Document Analyser can be improved by using more advanced AI models trained specifically on legal data to better understand complex legal terms and regulations. It can also support multiple languages to make the system useful globally. The application can be developed into a cloud-based platform with real-time collaboration features and predictive analytics to identify risks and suggest better contract terms. Integration with government databases and secure technologies like blockchain can further improve accuracy, transparency, and data security.

REFERENCES

1. A. Gupta, R. Sharma, and V. Arora, "AI-powered legal document summarization using NLP techniques," *IEEE Access*, vol. 12, pp. 45123–45134, 2024.
2. M. Lee and J. Kim, "Leveraging transformer-based models for contract clause identification," in *Proc. IEEE Int. Conf. on Artificial Intelligence and Law (ICAIL)*, Seoul, South Korea, pp. 89–96, 2023.
3. P. Singh, A. Mehta, and S. Bansal, "Intelligent legal assistant: An AI chatbot for automated contract analysis," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 765–774, 2024.
4. R. Zhao, L. Chen, and Y. Wang, "OCR and NLP integration for legal document analytics," in *Proc. IEEE Int. Conf. on Document Analysis and Recognition (ICDAR)*, San José, USA, pp. 302–309, 2023.
5. S. Patel and D. Nair, "Generative AI for question answering in domain-specific documents," *IEEE Intelligent Systems*, vol. 39, no. 1, pp. 56–64, 2024.
6. K. Tanaka and H. Suzuki, "Towards privacy-preserving AI in legal tech applications," *IEEE Access*, vol. 11, pp. 87965–87974, 2023.
7. Y. Choudhary, A. Bhatt, and N. Saxena, "Neural approaches for summarization of legal case documents," in *Proc. IEEE Int. Conf. on Natural Language Processing (ICONLP)*, Delhi, India, pp. 118–125, 2022.
8. F. Zhu, C. Liu, and X. Yang, "Document understanding with multimodal transformers for text and layout analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 2902–2914, 2023.
9. T. Ahmad, M. Farooq, and J. Alam, "AI-driven clause detection and semantic alignment in legal contracts," *Expert Systems with Applications*, vol. 226, p. 120225, 2023.
10. J. Park, E. Han, and M. Kim, "Fine-tuned language models for legal document summarization and entity recognition," *ACM Transactions on Asian and Low- Resource Language Information Processing (TALLIP)*, vol. 23, no. 1, pp. 1–15, 2024.