



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 12, December 2025

AI-Power Legal Document Analysis with an Integrated Trust and Safety Framework

Vivek Venkatesh Kulkarni, Mr Ajith GL

P.G. Student, Department of Master of Computer Application, PES Institute of Technology and Management,
Shivamogga, Karnataka, India

Assistant Professor, Department of Master of Computer Application, PES Institute of Technology and Management,
Shivamogga, Karnataka, India

ABSTRACT: Manually reviewing legal documents is time-consuming, costly, and prone to errors, which inspired the creation of the “Legal Analyzer,” a complete system built for accuracy and reliability. It uses advanced models to identify legal clauses and provide clear, actionable suggestions. To build user confidence, the system includes a Trust and Safety dashboard that allows auditing of fairness, evaluation of privacy risks, and assessment of prediction reliability. The main model performs effectively, achieving a weighted F1-Score of 0.849 on the CUAD dataset. It produces interactive reports that turn complex legal documents into easy-to-understand summaries, risk scores, alerts, and prioritized recommendations. Overall, the project offers a practical, transparent, and dependable solution for reviewing legal documents efficiently.

KEYWORDS: Legal document review, Clause classification, Risk scoring, Trust dashboard, Interactive reporting, Document summarization, BERT, NLP, safety

I. INTRODUCTION

This project introduces the Legal Analyzer, a comprehensive web application designed to simplify and improve the review of legal documents. It is built on a modern service-oriented architecture, featuring a React.js frontend, a Node.js API, and a background Python worker that handles complex processing while keeping the system responsive. Users can upload multiple documents at once, including PDFs, DOCX files, or scanned images. The system first cleans the documents by removing layout-based noise such as headers, footers, and stamps. When necessary, it extracts text from images using Optical Character Recognition (OCR). Important legal clauses are then identified, and a risk assessment module assigns scores to highlight potential issues. The system also generates clear, actionable recommendations to help reduce risks. A key feature is the Trust and Safety dashboard, which enables users to evaluate privacy using weighted scoring, check fairness with performance metrics, and view transparency through confidence indicators. Users can provide feedback to maintain accountability. The application produces interactive reports that summarize the document’s risk profile, highlight missing or overlooked clauses, and provide detailed explanations. This approach allows users to make faster and better-informed decisions. By combining automated clause detection, risk evaluation, and detailed reporting, the Legal Analyzer offers a practical, reliable, and user-friendly solution for reviewing legal documents.

II. LITERATURE SURVEY

A review of the current literature shows important achievements in Legal NLP. It also highlights the key gaps that this study aims to address. This study examines the connections between MLOps, AI ethics, and legal natural language processing. It highlights how today’s legal NLP models must be domain-specific.[1] This study explains the weaknesses of traditional keyword-based search used in legal research. The authors propose a multi-stage machine learning system that uses semantic features to recommend more relevant legal documents. Their approach handles the complex structure of legal text better than simple keyword matching, making document search smarter and more accurate.[2] This paper discusses how AI can improve legal processes in countries like India. It highlights the benefits of automating tasks such as legal research, document preparation, and verification. The authors also warn about important concerns like algorithmic bias and data privacy, emphasizing the need for safe and responsible AI.[3] This

research introduces InLegalBERT, a transformer model trained specifically on Indian legal documents. The study shows that domain-specific models outperform general models in tasks such as statute identification, segmentation, and judgment prediction. The results prove the importance of specialized training for legal applications.[4]The authors present a hybrid method using Law-T5 for summarizing complex contracts and Law-GPT for identifying risks from the summarized data. Their approach achieves an F1-Score of 0.80 on the CUAD dataset, showing the potential of combining summarization with risk detection[5]This study addresses the difficulty of interpreting ambiguous legal text. The proposed system combines a supervised BERT model with a Reinforcement Learning agent that adjusts predictions based on feedback. This helps the model adapt to different legal contexts and improves classification accuracy.[6]This research looks at summarizing legal documents in Turkish, a language with an agglutinative structure that makes analysis challenging. The authors compare several transformer models using ROUGE scores and find that GPT-3.5 Turbo performs the best among them. The study also highlights the need for language-specific optimization.

III. PROPOSED METHODOLOGY

The Legal Analyzer is an AI-enhanced web-based platform designed to deliver a complete, intelligent, and unified solution for legal document analysis. It integrates text extraction, clause classification, risk assessment, and generative insights into a single cohesive workflow. After secure authentication, users access a dashboard displaying analytics from previously analyzed documents, clause distribution summaries, and risk insights. The system processes uploaded contracts in formats such as PDF, DOCX, and scanned images, applying OCR where necessary to ensure accurate text retrieval. The clause classification module uses a fine-tuned BERT model to identify and categorize various legal clauses, while a GPT-based component provides contextual recommendations, explanation of risks, and suggestions for improvement. Before generating insights, users simply upload their document, and the system automatically analyzes structure, clause types, and risk indicators. Interactive risk reports help users understand document health by highlighting strengths, gaps, and compliance concerns. The platform also features an AI-based privacy module that detects and redacts sensitive information (PII), enhancing safety during analysis. Its integrated design ensures that clause detection, risk evaluation, and explanation work together, eliminating the need for separate tools. The interface is responsive across different devices, and strong authentication ensures secure access. The system is built for scalability and future expansion, supporting additional document types, legal standards, and advanced analytic

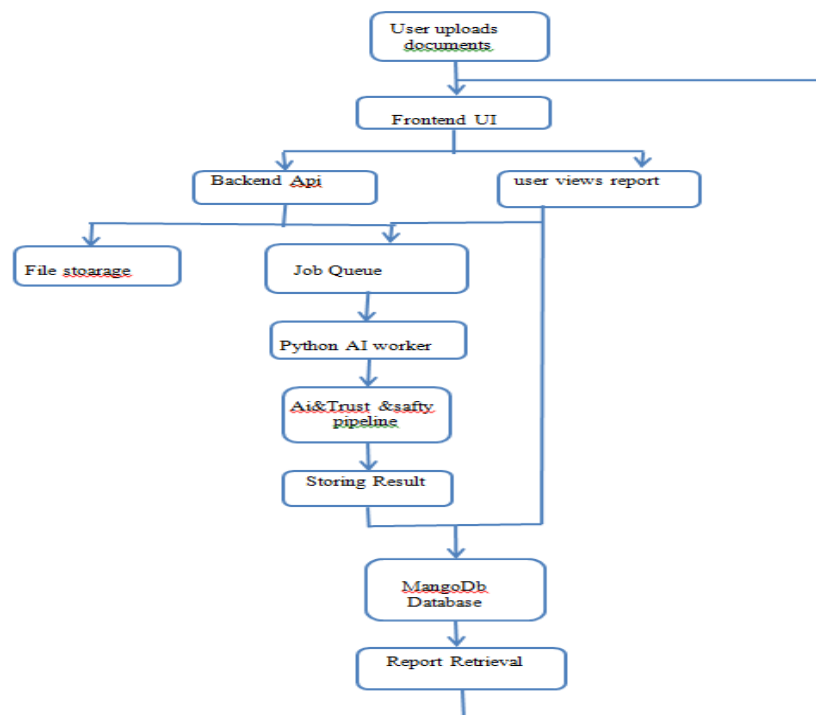


Fig. 1 Data Flow Diagram

The data flow diagram in Fig. 1 illustrates the end-to-end workflow of the system. Users begin by uploading documents through the frontend UI, which communicates with the backend API to handle requests. Once uploaded, files are stored in the file storage and queued for processing via the job queue. A Python AI worker then processes the documents, integrating components for AI, trust, and safety to ensure secure and reliable analysis. The processed results are stored in the MongoDB database, forming the basis for report retrieval. Users can then view the generated reports through the frontend interface, completing the cycle. This pipeline ensures an organized, efficient, and secure handling of documents from upload to final report delivery.

Algorithms Used: The Legal Document Analyzer uses different AI models to read documents, find risks, and give clear guidance. BERT identifies important legal clauses, while GPT creates easy suggestions for any issues found. spaCy's NER hides personal details like names and dates, and a risk-scoring system checks keywords to rate the document from 0 to 10. The system also cleans the text and uses OCR for scanned files. Finally, the results are shown on the dashboard with key metrics like Precision, Recall, and F1-Score to ensure accuracy.

1. **BERT (Bidirectional Encoder Representations from Transformers):** BERT (Bidirectional Encoder Representations from Transformers) is a powerful deep-learning language model that understands text by reading it in *both directions* at the same time, allowing it to capture the full context and meaning of complex legal language. In a Legal Document Analyzer, BERT is especially valuable because legal texts often contain long sentences, layered clauses, and terms whose meaning depends heavily on surrounding context. By leveraging its bidirectional attention mechanism, BERT can accurately identify key legal entities, interpret obligations and conditions, detect inconsistencies, classify document types, and extract important clauses such as confidentiality, indemnity, or termination terms. Unlike older models that process text from left to right, BERT's contextual understanding enables it to grasp nuanced legal phrasing, making the system more reliable for tasks like contract review, compliance checking, risk identification, and summarization. Overall, BERT serves as the core intelligence that allows the Legal Document Analyzer to interpret legal documents with higher precision and human-like understanding.

2. **GPT (Generative Pre-trained Transformer):** GPT (Generative Pre-trained Transformer) is an advanced AI language model designed to understand and generate human-like text. It is trained on massive amounts of data, allowing it to predict the next word in a sentence and create meaningful, context-aware responses. GPT can perform tasks such as answering questions, summarizing documents, generating ideas, rewriting content, and providing suggestions with high accuracy. Because of its ability to understand context and produce natural-sounding text, it is widely used in applications like chatbots, content creation, coding assistance, and smart document analysis.

3. **Weighted Risk Scoring :** A Weighted Risk Scoring is a method used to evaluate the risk level of a document by assigning different weights to specific keywords or patterns. Each keyword represents a potential risk, and the system gives higher importance (weight) to more critical terms. As the document is analyzed, the algorithm adds up these weighted values to calculate a final score, usually on a fixed scale such as 0 to 10. This score helps users quickly understand how risky a document is and decide whether further review or revision is needed.

4. **Tesseract OCR:** Tesseract OCR (Optical Character Recognition) is an open-source tool used to convert scanned documents or images containing text into editable and searchable digital text. It works by analyzing the shapes of characters in an image and matching them to actual letters, words, and numbers. This makes it useful for processing printed or handwritten documents, PDFs, or photos, allowing systems to extract text automatically. Tesseract is widely used because it supports many languages and integrates easily into applications that require text recognition.

5. **Layout-Aware Noise Filtering:** Layout-Aware Noise Filtering is a method used to clean document text while still keeping its original structure intact. Instead of deleting content randomly, it studies the layout like headers, footers, page numbers, watermarks, and repeated patterns to decide what should stay and what should be removed. By understanding how the document is arranged visually, it removes unwanted elements without disturbing the main text. This results in cleaner, clearer content that is easier for analysis, especially when dealing with scanned or complex documents.

Testing: System testing for the Legal Document Analyzer ensures that all integrated components—such as document upload, backend processing, AI analysis, trust and safety checks, and report generation—work together correctly as a complete system. During this phase, the entire workflow is tested end-to-end to verify that documents are processed accurately, results are stored properly, user interactions function smoothly, and the system behaves reliably under different conditions. The goal of system testing is to confirm that the analyzer meets all functional requirements, handles errors gracefully, maintains data security, and delivers a consistent user experience before deployment.

Various types of Software testing

1. Unit Testing: Unit testing for the Legal Document Analyzer focuses on validating each individual component of the system in isolation, ensuring that every module functions correctly on its own before integration. This includes testing the document upload handler, API endpoints, AI processing functions, job queue operations, database interactions, and trust & safety utility functions. Each unit is tested with both valid and invalid inputs to verify accuracy, stability, and error handling. By identifying defects early at the module level, unit testing helps improve overall system reliability and ensures that the foundation of the analyzer is robust before combining components into the full workflow.

2. Integration Testing: Integration testing for the Legal Document Analyzer ensures that individual modules work correctly when combined, focusing on the interactions between components such as the frontend, backend API, job queue, AI processing pipeline, and database. This testing verifies that data flows smoothly from document upload to analysis, that the AI worker receives jobs properly, that results are stored without errors, and that the frontend can fetch and display reports accurately. By checking the communication and interfaces between system parts, integration testing helps detect issues that may not appear during unit testing and ensures the overall workflow functions cohesively..

3. Validation Testing: Validation testing for the Legal Document Analyzer ensures that the final system meets all user requirements and performs the intended functions accurately. This testing focuses on confirming that features such as document upload, AI-based analysis, trust and safety checks, report generation, and feedback handling operate according to the specifications defined in the project scope. It ensures that the system delivers correct results, maintains data security, and provides a smooth user experience. By comparing actual performance with expected outcomes, validation testing verifies that the system is ready for real-world use and satisfies the needs of legal professionals.

4. System Testing: System testing for the Legal Document Analyzer focuses on evaluating the entire application as a complete and fully integrated system. It checks the full workflow—from uploading documents, processing them through the AI pipeline, performing trust and safety checks, storing results, and finally presenting the report to the user. This testing ensures that all modules work together smoothly, that data moves correctly between components, and that the system behaves reliably under normal and unexpected conditions. By validating performance, security, and overall functionality, system testing confirms that the Legal Document Analyzer is stable, meets project requirements, and is ready for deployment.

5. Acceptance Testing: Acceptance testing for the Legal Document Analyzer is performed to ensure that the system meets the expectations and requirements of the end users before final deployment. This testing focuses on real-world usage scenarios such as uploading legal documents, running the AI analysis, performing trust and safety checks, and viewing the generated report. It verifies that all key features are functional, the interface is easy to use, and the system delivers accurate and meaningful results. By validating the system's usability, correctness, and readiness, acceptance testing confirms that the Legal Document Analyzer is suitable for practical use by legal professionals.

Test Cases:

Test Case ID	Module	Test Objective	Test Steps	Expected Result	Status
TC-01	Document Upload	Verify successful upload of multiple, mixed-format documents.	1. Navigate to the Upload page. 2. Select a PDF, DOCX, and PNG file simultaneously. 3. Submit the batch for analysis.	System accepts all files, creates a single analysis job, and the job status appears as "Pending" on the dashboard.	Pass
TC-02	Document Upload	Verify rejection of invalid file formats.	1. Navigate to the Upload page. 2. Attempt to upload an unsupported file type (e.g., .zip).	System displays an error message indicating the file type is not supported and rejects the upload.	Pass
TC-03	AI Analysis Pipeline	Verify the Layout-Aware Noise Filtering.	1. Upload a scanned document (PDF/PNG) with a prominent e-Stamp	Text from the e-Stamp header is not present in analyzed clauses or executive summary.	Pass

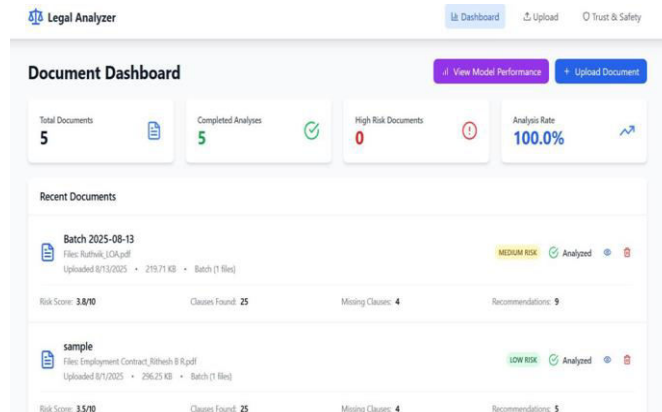
Test Case ID	Module	Test Objective	Test Steps	Expected Result	Status
			header. 2. Await analysis completion. 3. Review the final report.		
TC-04	AI Analysis Pipeline	Verify the core Clause Classification functionality.	1. Upload a standard contract with "Termination" and "Liability" clauses. 2. Review the final report.	Detailed Clause Analysis correctly identifies and groups clauses under "Termination Clauses" and "Liability Clauses."	Pass
TC-05	AI Analysis Pipeline	Verify the Risk Scoring and Tiered Suggestion feature.	1. In the detailed analysis, locate a clause with high-risk language (e.g., "sole discretion").	Clause is assigned high risk (>8/10), flagged as "CRITICAL," and AI Suggestion begins with "HIGH RISK – PRIORITY NEGOTIATION NEEDED."	Pass
TC-06	AI Analysis Pipeline	Verify the Missing Standard Clauses detection feature.	1. Upload a document missing a "Dispute Resolution" clause. 2. Review the final report.	"Missing Standard Clauses" section lists "Dispute Resolution" as missing.	Pass
TC-07	Trust & Safety	Verify the PII Redaction and Privacy Score.	1. Navigate to Trust & Safety dashboard. 2. Enter text containing email and SSN into PII Redaction Tool. 3. Click "Redact PII."	Text is redacted, and a low "POOR" or "CRITICAL" privacy score is generated with detailed breakdown.	Pass
TC-08	Trust & Safety	Verify the in-app Model Evaluation Suite.	1. Navigate to Model Performance Evaluation modal. 2. Click "Run New Evaluation."	Displays Overview, Detailed Metrics, and Confusion Matrix with correct pre-validated scores (Accuracy: 84.7%, F1: 0.849).	Pass
TC-09	Trust & Safety	Verify Human Feedback submission system.	1. Navigate to Human Feedback tab. 2. Input correction for model prediction. 3. Click "Submit Feedback."	"Feedback submitted successfully!" message appears and dashboard counter increments.	Pass
TC-10	Final Report	Verify "Export Report" functionality.	1. On a completed analysis page, click "Export Report."	Formatted report (PDF) with executive summary, charts, and detailed clause analysis is successfully generated and downloaded.	Pass

IV. EXPERIMENTAL RESULTS

The image shows the Legal Document Analyzer dashboard, where users can access tools like the Document Upload, AI Analysis Pipeline, Trust & Safety Suite, and Report Export. The dashboard also provides quick updates on recent analyses, job statuses, and key insights from processed documents, enabling users to efficiently monitor and manage their legal review workflow.

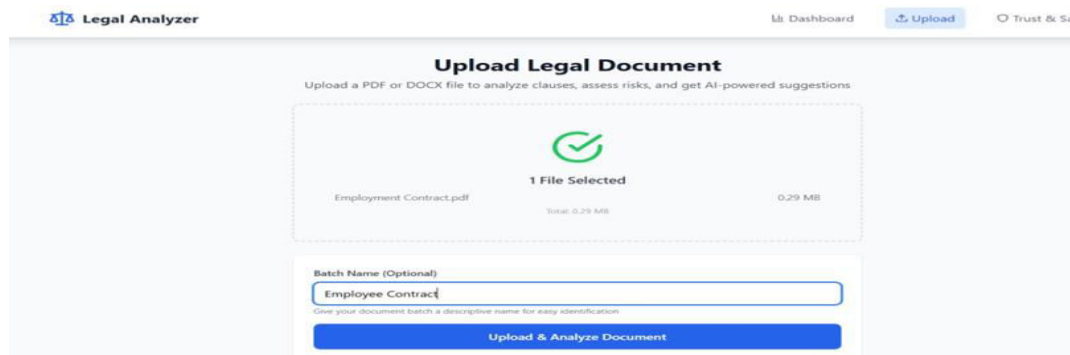
Volume 14, Issue 12, December 2025

[DOI: 10.15680/IJIRSET.2025.1412049]



THE MAIN DOCUMENT DASHBOARD INTERFACE

Document Dashboard: A central hub showing past analyses, risk levels, and upload options, with quick access to the Trust & Safety section for model performance



The Document Upload and Batch Naming Interface.

This screen shows the first step in the analysis workflow, where users can select one or more legal documents in various formats (PDF, DOCX, etc.). The system allows users to group related files into a single "batch" and assign a descriptive name, ensuring all documents for a specific case are analyzed together

V. CONCLUSION

This project was undertaken to address the dual challenges of inefficiency and the lack of integrated ethical governance in modern legal technology. To this end, a comprehensive system, the "Legal Analyzer," was successfully developed, integrating a high-performance, fine-tuned BERT model for clause classification with a generative GPT model for producing actionable suggestions. The evaluation of the analytical core demonstrated strong quantitative performance, achieving a weighted F1-Score of 0.849 and an overall accuracy of 84.7%, validating the effectiveness of the chosen approach.

The key innovation of this research was the successful implementation of a complete Trust & Safety framework and an in-application evaluation suite, which directly address critical gaps in the current legal technology landscape concerning transparency, accountability, and verifiable performance. The successful implementation of the interactive dashboards for privacy, fairness, and model evaluation demonstrates a practical methodology for operationalizing the principles of Responsible AI. Ultimately, this project serves as a working model for a new generation of legal AI tools, proving that high analytical performance and robust safety measures are not mutually exclusive but are, in fact, synergistic components of an advanced and trustworthy system.

Volume 14, Issue 12, December 2025**|DOI: 10.15680/IJIRSET.2025.1412049|****REFERENCES**

- [1] Andrews, M., Bromiley, P., Chow, E., & Gibson, T. (2024). A Machine Learning Framework for Legal Document Recommendations. *Journal of Computer Science and Artificial Intelligence*, 1(1).
- [2] Nithya, M., Harini S, Kavyadharshini S, & Srinidhi K. (2024). AI-Driven Legal Automation to Enhance Legal Processes with Natural Language Processing. *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*.
- [3] Paul, S., Mandal, A., Goyal, P., & Ghosh, S. (2022). Pre-training Transformers on Indian Legal Text. *arXiv preprint arXiv:2209.06049*.
- [4] Poornima A, Nagaraja K.V, & Venugopalan, M. (2025). Legal Contract Analysis and Risk Assessment using Pre-Trained Legal-T5 and Law-GPT. *2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*.
- [5] Bhushan, S., Hemaditya AVS, & Verma, A. (2025). Hybrid Legal Clause Classification: Integrating Supervised BERT with Reinforcement Learning. *2025 3rd International Conference on Disruptive Technologies (ICDT)*.
- [6] Arun Kumar R, Sairam N, Aswanth VR, Saikrishna V, & Palanikumar R. (2025). Revolutionizing Legal Workflows: Advanced AI Techniques for Document Summarization, Legal Translation, and Conversational Assistance. *2025 International Conference on Advanced Computing Technologies (ICOACT)*.
- [7] Albayati, M. A. A., & Findik, O. (2025). A Hybrid Transformer-Based Framework for Multi- Document Summarization of Turkish Legal Documents. *IEEE Access*, 13.
- [8] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. *arXiv preprint arXiv:2010.02559*.
- [9] Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 conference on fairness, accountability, and transparency*.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details