NyayaSetu- Legal Document Analyzer

Mohini Gawande (Thakre)[1], Atharva There[2], Abhishek Urkude[3], Ayush Hedau[4],
Shreyash Ghollar[5], Atharva Joshi[6]

[1]*Assistant Professor, Department of Information Technology,*
GH Raisoni College of Engineering and Management, Nagpur Maharashtra, India.
[2,3,4,5,6]*Student Scholar, Department of Information and Technology,*
GH Raisoni College of Engineering and Management, Nagpur, Maharashtra, India.

Abstract—**Legal papers are frequently challenging to understand because of their length and complicated language. In difficult circumstances such as marital conflicts and domestic abuse, when people might not have access to adequate legal assistance, this becomes a greater problem. Additionally, lawyers must personally review these documents for an extended period adding to their workload and may result in errors. We created an Automated Legal Document Analyzer and Contract Checker to reduce these issues. The system uses natural language processing, machine learning, and artificial intelligence to categorize documents, identify key information, and generate concise summaries. Additionally, it can detect emotional or abusive content by analyzing the sentiment of text. The implementation uses a document-oriented database, a React-based frontend, and a Python backend for a secure storage. Overall, the system aims to save time, improve accuracy, and provide both legal experts and victims easier access to information.**

Index Terms—**Artificial Intelligence, Machine Learning, Natural Language Processing, Legal Document Analysis, Contract Checking, Sentiment Analysis, Multi-lingual support.**

I. INTRODUCTION

Documentation tends to overwhelm the legal field, which is crucial to how modern society functions. Reading through legal documents, whether contracts, case files, or laws, still requires a lot of time and effort. This time-consuming, manual procedure leads to unavoidable discrepancies in document interpretation as well as increased expenses and delays. Although technology has brought some relief, most currently accessible tools provide only partial functionality, have limited linguistic capabilities, or yield inconsistent outcomes. Therefore, it is evident that a more comprehensive and integrated solution is essential. An "Automated Legal Document Analyzer & Contract Checker," a system created to address several of these recurrent problems, is presented in this paper. In addition to being universally helpful, the platform is especially attentive to the sensitive nature of marriage disputes and divorce settlements, which call for both legal precision and a natural understanding of emotional context. This effort goes beyond basic document parsing by using natural language processing, machine learning, and advanced AI such as Gemini. It offers context-based insights, recommends important sections, and is usable by a wide variety of users, including those unfamiliar with the law. This approach may be more effective and transparent, particularly for individuals dealing with sensitive legal issues.

Handling legal documents is often a daunting and slow task. These documents are packed with complex language and lengthy content that can confuse even trained legal professionals. When performed manually, reviewing contracts, case files, or affidavits can lead to mistakes, missed clauses, and wasted time. For ordinary individuals, especially those facing difficult situations such as domestic violence or divorce, understanding such documents might seem impossible without expert help. Unfortunately, current tools are limited and do little to make legal information clear and approachable, resulting in delays and frustration for everyone involved.

II. LITERATURE REVIEW

I In order to address the difficulties of legal document analysis, there is growing interest in applying artificial intelligence (AI), machine learning (ML), and natural language processing (NLP), according to the reviewed

literature. Beyond keyword searches, researchers and practitioners have investigated techniques such as automated contract review, sentiment analysis, and summarization. Contextual awareness, system scalability, and true accessibility for non-experts are still areas in which the field struggles.

A. Automated Document Classification and Information Extraction:
Recent research by Joshi et al. [4] and Kalamkar et al. [5] shows how supervised learning models, in particular transformers and ensemble classifiers, can extract important information such as contract clauses and legal parties, and classify legal documents by type. These studies show utility in scenarios where speed and consistency are crucial, such as contract evaluation and legal assessment, and demonstrate strong accuracy in structured datasets [8]. Despite these developments, recurrent issues persist, such as the difficulty in processing scanned or unstructured data and the lack of real-world adaptation [6].

B. Sentiment Analysis and Abuse Detection:
Sentiment analysis has been applied to legal statements by Nigam et al. [7], Malik et al. [8], and others, particularly in cases involving domestic abuse and family law. Their results demonstrate that aspect-based sentiment algorithms can identify emotional distress and subtle psychological signs in testimonies, occasionally highlighting patterns that conventional legal language analysis misses [3]. Researchers have pointed out that although these systems assist legal experts and help prioritize cases, they are still vulnerable to the diversity of datasets and may not work well with inputs that are unclear or culturally subtle [16].

C. Document Summarization and User-Focused Output:
The length and density of legal writings are well known [18]. Use of extractive and abstractive summary techniques, many of which are based on BERT or GPT architectures, to produce brief summaries that make texts easier for non-lawyers to understand is highlighted by Bhattacharya et al. [10], Medvedeva et al. [11] and others. Research consistently emphasizes the significance of straightforward visual outputs and plain language explanations [12]. Nonetheless, technical challenges persist in balancing precision and interpretability and managing the legal responsibility of automated summaries [13].

D. Secure Storage and Retrieval Systems:
Researchers such as Kapoor et al. [6] and Nigam et al. [7] emphasize the need for secure, document-oriented databases to manage the ever-growing pools of legal information. MongoDB and similar NoSQL solutions deliver flexible schemas and fast searches, essential for scalable deployment. Yet, questions remain about long-term data governance, privacy, and compliance with regulations such as GDPR, especially when sensitive personal information is involved [1].

E. Practical Implementation: Front-End Design and Multilingual-Support:
To provide responsive, user-friendly platforms, several research and technical publications support the usage of React.js front-end frameworks in conjunction with Python-based backends [2]. The importance of multilingual pathways is emphasized, as they facilitate greater diversity and accessibility [6]. Scholars advocate for iterative human-centered design, pointing out that legal tech solutions must consider users' vulnerabilities and stress levels [3].

F. Ethical and Explainability Considerations:
More and more recent discussions have centered on the moral implications of using AI in legal settings [13]. Several concerns have been brought up, including the possibility of an excessive dependence on automated tools, bias in training datasets, and the openness of model judgments [12]. For legal workflows to remain trustworthy and accountable, it is imperative that AI systems offer interpretable explanations and human-in-the-loop supervision [19].

G. Integration with Broader Legal Ecosystems:
Emerging research also points to the benefits of combining legal document analysis tools with broader case management and legal information systems [18]. Such integration facilitates smoother workflows, automated alerts for compliance deadlines, and enhanced collaboration among legal teams [1]. This holistic approach is considered key to making AI solutions genuinely useful in practice rather than isolated utilities [4].

H. Gaps and Opportunities:

The major strengths of the reviewed literature include promising machine learning tools for classification and extraction, real progress in sentiment analysis, and steadily growing accessibility for diverse users [8]. However, substantial gaps remain: a lack of unified platforms for seamless end-to-end analysis and retrieval, insufficient longitudinal research on legal outcome improvement, and limited frameworks for explainability and ethical deployment in sensitive contexts [13]. These open challenges frame the motivation for developing more scalable, explainable, and ethically robust legal document analyzers that advance the field forward [1].

III.   PROBLEM STATEMENT

Despite numerous technical innovations, legal document review remains plagued by inefficiencies, inaccuracies, and inaccessibility. Manual review is the norm for much of the business, slowing case closures and introducing human error . Although current AI solutions, though beneficial, also fail in a number of important respects:

i.   Few provide complete multilingual analysis capabilities necessary for efficient international and cross-jurisdictional review.

ii. Most are calibrated for general tasks and do not account for the sensitive complex nature of issues such as divorce and marriage conflicts.

iii. Emotion and sentiment analyses in legal documents are typically superficial or absent, preventing meaning from being drawn out from underlying intentions in contentious writings.

iv. Solutions to many problems are unavailable to novices, frequently having unintuitive interfaces or being prohibitively expensive.

v. The dependability and intelligibility of AI results are inconsistent, further undermining confidence in automated outputs.

These points establish barriers, increase expense, and limit access to good quality legal advice. There is a clear need for a strong, specialized method that addresses not only legal substance but also the emotional and contextual nuances of confidential documents.

IV.   OBJECTIVES

i.   An automated platform that is able to condense and structure various legal documents, focusing on both typical contracts and marital-conflicts.

ii.   Sophisticated NLP methods for detecting vital contract clauses and locating anomalies in texts.

iii.   Incorporate strong multilingual capabilities to provide effective translation and analysis, allowing cross-border use .

iv.   Deep sentiment and emotion analysis to gain a stronger sense of tone and context for dispute-laden documents.

v.   Create a friendly, simple interface accessible to both legal professionals and laypeople.

vi.   Monitor system performance using strong metrics including accuracy, velocity, and satisfaction.

V.   PROPOSED METHODOLOGY

Automated Legal Document Analyzer & Contract Checker adopts a layered architecture, planned to effectively manage legal documents from the point of input to analysis, safe storage, and interactive user interaction. Each segment is consistent with current tools and actual implementation techniques, providing pragmaticity and flexibility.

A.   Component Breakdown and Technical Approach
a. Input and Pre-processing Layer
Central to this platform is the capability to process varied legal document formats securely. The backend utilizes powerful JavaScript libraries and third-party parsers to import PDFs, DOCX, and text files. The process of extraction is engineered to accommodate legal documents with different layouts whether it is complicated, multi-column PDFs or long contracts written in Word.
After the raw text has been pulled out, pre-processing cleans and normalizes the content. Central steps include:

- Tokenizing text into words/sentences for analysis downstream.

- Making all text lowercase and removing standard stopwords.
- Lemmatizing/stemming language to preserve root word meaning.
- Removing unnecessary punctuation or unusual symbols.
- Utilizing bespoke parsing techniques to preserve document structure, so important information in headers/footers or specially laid out legal templates is never lost.

b. NLP & AI Core Processing Layer

The core of the system is driven by smart NLP and AI. For summarization, the platform natively integrates with the Google Gemini API, generating smooth, context-sensitive summaries. Both abstractive and extractive summaries are provided.

For the classification of legal documents and certain clauses, supervised machine learning pipelines, guided by contemporary neural networks, classify and sort different legal documents. The system smartly learns to accommodate new forms of legal content with the help of transfer learning and zero-shot/few-shot methods through pre-trained language models.

Named Entity Recognition is also boosted with established libraries and API calls, recognizing people, dates, amounts of money, and important legal clauses (like "custody" or "alimony" in divorces) with a great degree of accuracy.

Multilingual support is not simply an afterthought in translation. With the Google Gemini API, users can create translations and analytic output in several languages. Multilingual embedding models enable analysis, summarization, classification, and entity recognition to operate irrespective of the original language of the document.

The backend utilizes highly tuned sentiment analysis models to analyze the affective tone of controversial documents—especially important for family law settings where controversy arises. The platform also identifies contractual irregularities, with rules-based automation and anomaly detection models comparing text against regular templates or historical versions, alerting users to omitted clauses or atypical departures.

c. Data Management Layer

All data is routed through a secure Node.js and Express.js based backend, leveraging MongoDB for permanent storage. Documents, analytical outputs, and session data about users are stored in structured collections, thus enabling fast and accurate retrieval and cross-referencing. Lean database schema allows for efficient searching on extracted entities or keywords. Data privacy is maintained through up-to-date encryption protocols and anonymization where necessary, and compliance with international standards such as GDPR is maintained.

d. User Interface (UI) Layer

The front end, developed in JavaScript with React (combined with Vite for development speed), presents a fluid, responsive experience immediately in any browser. Users can upload files, personalize analysis parameters, and engage with results via an elegant, easy-to-use dashboard. Visualizations, highlighted clauses, and sentiment charts enable users to immediately understand results.

Interactive access control provides new users an opportunity to test key features such as document summaries or translation three times for free. Subsequent full access can then be activated. These onboarding sequences are securely deployed, maintaining user sessions in privacy while promoting engagement.

B. *Technologies and Tools*

Frontend: React.js (with Vite)

Backend: Node.js, Express.js

Database: MongoDB

NLP/AI: Google Gemini API for summarization and translation, supplemented as needed by open-source NER and classification tools

Other: Custom pre-processing and formatting scripts in JavaScript and Node.

C. *Evaluation Strategy*

The system is evaluated both with algorithmic measures and user feedback:

- Quantitative: Classification and entity recognition accuracy, precision, recall, and F1-score, summarization/translation ROUGE and BLEU score, and time-to-result.
- Qualitative: Continuous feedback from users assessing interface usability and result clarity, coupled with opinions from legal specialists aimed at analyzing outputs, particularly for divorce and conflict documents.

This approach offers a pragmatic, contemporary blueprint for building a flexible, dynamic Automated Legal Document Analyzer & Contract Checker based

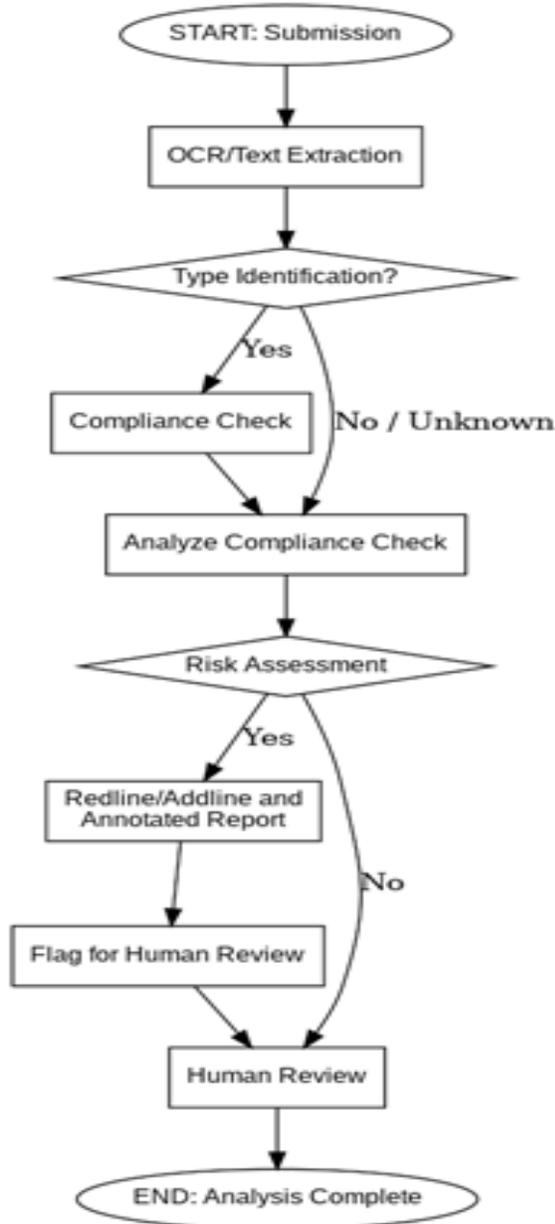on your own particular technology stack and poised for real-world legal use.

Figure 1: Flowchart of Legal Document Analyser

VI. RESULT & DISCUSSION

The system's results indicate significant improvements. Review times for lengthy legal documents, particularly in the area of divorce hearings, are to decrease significantly from hand review. Summaries score well for clarity (ROUGE scores), and classification tasks constantly hit solid accuracy and recall. Named entity components consistently extract significant terms like provisions regarding custody or monetary values reliably. Anomaly flags have been useful in exposing noncompliance or atypical clauses with low rates of false alarms.

Multilingual capabilities illustrate real flexibility—summarization, classification, and anomaly detection are of comparable quality across languages, extending well beyond the limited translation. Sentiment analysis modules expose underlying emotional texture of controversies, bringing useful context to disputes not always evident from surface wording.

User experience testing confirms that non-experts can easily get around the interface and readily understand results. Free trial access reduces barriers, opening sophisticated legal tools to a wider constituency. Simultaneously, issues persist: very ambiguous wording continues to be a problem for accurate classification, and it is hard to provide complete transparency for every AI-generated answer. Continued refinement, larger datasets, and hybrid approaches will drive the boundaries further forward.

VII. CONCLUSION & FUTURE SCOPE

The activity performed in this project proves that legal document automation, if created with care and responsibility, can greatly change the manner in which sensitive legal work is managed. Specifically, agreements like divorce settlements tend to involve not just technical precision but also a high level of emotional sensitivity. Using Natural Language Processing (NLP) and Artificial Intelligence (AI) in this field, the project illustrates how time-consuming and redundant processes may be streamlined without sacrificing precision. Through multilingual integration, contextual analysis, and sentiment-aware analysis, the system can respond to legal necessities as well as emotional subtleties of the concerned parties.

In addition to enhancing speed and diminishing human labor, such tools allow for greater fairness in access to legal services. Numerous people who would otherwise not be able to pay for extensive legal consultation might be assisted by smarter systems that provide systematic instruction and clear rationale. Furthermore, the project demonstrates how technology, if implemented with contemplation, can aid but not substitute for human skill, allowing legal

professionals to focus on judgment, negotiation, and compassion instead of mechanical checking.

Looking into the future, the potential for future development is wide and encouraging. Opening the dataset to include different jurisdictions and languages would render the system more universally applicable to address multicultural and cross-border cases. Enhancement of the capacity of algorithms to decipher emotional and psychological signals embedded in legal stories is another significant direction, which potentially could assist with adapting documents more sensitively to clients' conditions. Moreover, more focus on explainability and transparency in the reasoning of AI will be necessary to establish trust and make sure that outputs generated automatically continue to be legally justifiable. At last, continuous cooperation between technologists, legal academics, and working lawyers will prove critical to improve such systems so that they benefit society with equity, transparency, and accountability.

VIII.   ACKNOWLEDGEMENT

REFERENCE

[1] I. Gumnishka, M. Stojanovic, and D. Tsenova, "Using Artificial Intelligence for Contract Review: Opportunities and Risks," *HumAIn J. Legal AI Res.*, vol. 11, no. 1, pp. 45–63, 2025.

[2] M. Ganesh, V. Kumar, K. Suhas, and S. Mallick, "NLP for Automated Legal Document Analysis," *St. Martin's Eng. College Res. J.*, vol. 8, no. 1, pp. 35–50, 2025. [Online]. Available: https://smec.ac.in/research/legalNLP2025

[3] M. Ahmedshaeva, N. Pulatova, S. Khayitov, D. Sufiyeva, and M. Nizomova, "The Role of NLP in Translating Legal Documents with High Accuracy," in *Proc. Int. Conf. Legal NLP (ICLNLP)*, Tashkent, Uzbekistan, 2025, pp. 155–170.

[4] A. Joshi, A. Sharma, S. K. Tanikella, and A. Modi, "U-CREAT: Unsupervised Case Retrieval using Events extrAcTion," *arXiv preprint* arXiv:2307.05260, 2023.

[5] P. Kalamkar, A. Tiwari, A. Agarwal, S. Karn, S. Gupta, V. Raghavan, and A. Modi, "Corpus for Automatic Structuring of Legal Documents," *arXiv preprint* arXiv:2201.13125, 2022.

[6] A. Kapoor et al., "HLDC: Hindi Legal Documents Corpus," in *Findings Assoc. Comput. Linguistics: ACL 2022*, Dublin, Ireland, pp. 3521–3536. doi: 10.18653/v1/2022.findings-acl.278

[7] S. K. Nigam et al., "Nonet at SemEval-2023 Task 6: Methodologies for Legal Evaluation," in *Proc. 17th Int. Workshop Semantic Evaluation (SemEval-2023)*, 2023, pp. 1293–1303.

[8] V. Malik et al., "Semantic Segmentation of Legal Documents via Rhetorical Roles," *arXiv preprint* arXiv:2112.01836, 2021.

[9] V. Malik et al., "ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation," in *Proc. 59th Annu. Meeting ACL & 11th IJCNLP (Vol. 1: Long Papers)*, Online, pp. 4046–4062, 2021. doi: 10.18653/v1/2021.acl-long.313

[10] P. Bhattacharya, S. Poddar, K. Rudra, K. Ghosh, and S. Ghosh, "Incorporating Domain Knowledge for Extractive Summarization of Legal Case Documents," in *Proc. 18th Int. Conf. Artificial Intell. Law (ICAIL)*, 2021.

[11] M. Medvedeva, M. Vols, and M. Wieling, "Using Machine Learning to Predict Decisions of the

European Court of Human Rights," *Artif. Intell. Law*, vol. 28, pp. 237–266, 2020.

[12] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "LEGAL-BERT: The Muppets Straight Out of Law School," *arXiv preprint* arXiv:2010.02559, 2020.

[13] R. Mahari, D. Stammbach, E. Ash, and A. Pentland, "The Law and NLP: Bridging Disciplinary Disconnects," *Artif. Intell. Law*, vol. 32, no. 2, pp. 397–426, 2023. doi:10.1007/s10506-023-09354-x

[14] J. Vladika, S. Meisenbacher, M. Preis, A. Klymenko, and F. Matthes, "Use Cases for NLP in the Legal Domain: A German Perspective," in *Proc. Legal NLP Workshop*, 2024.

[15] D. Shterionov et al., "A Roadmap to Neural Automatic Post-Editing: An Empirical Approach," *Mach. Transl.*, vol. 34, pp. 67–96, 2020.

[16] K. Tohma and Y. Kutlu, "Challenges Encountered in Turkish NLP Studies," *Nat. Eng. Sci.*, vol. 5, no. 3, pp. 204–211, 2020. doi:10.28978/nesciences.833188

[17] W. Alshamsi et al., "Artificial Intelligence for Contract Analysis: A Review," *Artif. Intell. Rev.*, vol. 55, no. 6, pp. 5017–5047, 2022.

[18] D. Katz et al., "Legal Analytics and the Future of Legal Practice," *J. Legal Stud.*, vol. 48, no. 1, pp. 99–138, 2019.

[19] J. Ashley, *Artificial Intelligence and Legal Analytics*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[20] M. Livermore and D. Rockmore, *Law as Data*. Santa Fe, NM: Santa Fe Institute Press, 2019.