

AI-Driven Legal Assistance Framework for Enhancing Judicial Efficiency and Legal Education in India

Jhalak Dedhia
Dept. of AI & DS KJSIT,
Mumbai, India

Dr. Sheetal Jagtap
Guide, Dept. of AI & DS
KJSIT, Mumbai, India

Heer Panchal
Dept. of AI & DS
KJSIT, Mumbai, India

Kinjal Panchal
Dept. of AI & DS
KJSIT, Mumbai, India

Abstract—The legal system of India faces very hard and huge challenges such as excessive case backlogs, time-consuming legal research and limited accessibility to legal resources across the country. This research represents an AI-driven legal assistance platform is designed to support the legal document summarization, legal document classifier, case law recommendation, virtual moot court. The system integrates the Natural Language Processing, Retrieval Augmented Generation and Speech Recognition techniques to automate the repetitive legal tasks. The platform aims to improve the efficiency, availability and learning outcomes inside the Indian legal system. The innovative evaluations shows improvements in document comprehension time, retrieval accuracy and educational effectiveness, highlighting the system's possibility to modernize legal workflows in an ethical and scalable manner.

Index Terms—Artificial Intelligence in Law, Courtroom Transcription, Indian Judiciary, Legal Education, Legal NLP, RAG Systems

I. INTRODUCTION

The Indian judicial system handles a huge number of cases every year, which often leads to delays in the judiciary system. Even though online platforms such as online case filing and e-Courts have been introduced, many legal activities still performed are done manually. Legal professionals spend a significant amount of time reading lengthy documents, searching for relevant case laws and preparing legal drafts. The manual processes also increase the workload and slow down overall judicial efficiency [4], [5].

Judgments, FIRs, and petitions all the documents are usually long and are frequently written in very hard legal language, which makes it difficult to understand. Law students and junior advocates also face challenges due to limited exposure to practical legal tasks such as case analysis and courtroom procedures. As a result, there is a growing need for intelligent systems that can assist legal professionals and learners in handling large volumes of legal information efficiently [1]–[3].

Artificial Intelligence (AI), especially in Natural Language Processing (NLP) and speech recognition has made it possible

to process legal text more effectively. The document summarization, classification, and automated speech-to-text transcription all these techniques are used to decrease the repetitive work and also maintaining human control over critical decisions [6], [9]–[11].

The paper used real world dataset that was provided by a professional advocate of High Court. The dataset consisted of about approximately 15,000 documents of judgments, pleadings, proceedings, case laws and orders from different areas of legal practice. Using such dataset helped enhance the quality and results of the project as the models performed well on real time and real world documents.

This paper suggests an AI-powered legal assistant focused on summarization, classification, legal research, courtroom transcription and legal schooling in the Indian context [7], [8], [12]–[14].

II. SURVEY OF RELATED WORK

There are many research that have contributed to the foundation of this project. Such research include legal text summarization, retrieval augmented generation (RAG), language models that are domain specific, knowledge graph based reasoning and automatic speech recognition (ASR) and many other research have contributed to the start of this project. The survey of related work section of this study provides a combined overview of existing work on AI applications in the Indian legal domain. This section also highlights the key contributions, innovations and persisting gap between existing frameworks and the proposed solution that together motivates the development of the proposed solution.

A. Legal Document Summarization

Previous work includes LawSum by Parikh et al. that created a weakly supervised extractive summarizer which extracted sentences through rhetorical cues and section marks, another legal summarizer was introduced by DELSumm by

Bhattacharya et al. that segmented legal judgments into Facts, Issues, Arguments and Decisions (FIAD) to improve interpretability and factual retention. Comparative studies done by Sharma et al. showed that fine tuned transformer like Legal Pegasus and BART outperform generic summarizers. However, the existing datasets are small in size and lack factual faithfulness evaluation. The solution proposed by the authors plans to address these gaps by using scoring and semantic consistency checks [1]–[3].

B. RAG-Based Conversational Systems

The generative flexibility of Large Language Models (LLMs) and precision from search retrieval are combined together by Retrieval Augmented Generation (RAG). Existing systems like LawPal, CREA2 and ChatLaw have integrated dense retrieval, knowledge graphs and sentence embeddings to enhance the accuracy, reduce hallucinations and improve legal citation source. These systems faced challenges in generalizing multilingual and low resource Indian languages. The assistant that is proposed addresses these gaps by using IndicBERT encoded embeddings that are multilingual and region specific to support adaptive legal dialogues [4], [7], [8].

C. Case-Law Recommendation and Legal Retrieval

For legal research and decision support, it is essential to have efficient precedent retrieval. There are traditional BM25 lexical searches that miss to capture the semantic reasoning, leading to combining BM25 with Sentence Transformer embeddings that improve retrieval performance. But the lack of Indian standardized open benchmark is leading to limited reproducibility. The proposed project uses hybrid semantic embeddings, citation graphs and contextual similarity metrics, that develops a publicly available and annotated Indian legal benchmark corpus.

D. Document Classification and Attribute Extraction

Legal analytics is structured by classifying and categorizing the legal texts into different domains and attributes such as IPC sections or verdict polarity. Adhikary et al. proposed weak supervision with few shot learning to solve the issue of limited data availability. Also, InLegalBERT and Legatron have used Named Entity Recognition (NER) and classification tasks to improve model performance. The framework used in this study enhances reliability in legal analytics by integrating together tagging, rule based post processing and citation linked validation [6], [9].

E. Courtroom Transcription and Diarization

The advancements of Automatic Speech Recognition (ASR) and speaker diarization have improved legal transcription. Models such as Whisper handles noisy and multilingual audio, whereas Pyannote provides improved speaker segmentation. Courtroom audio remains challenging as there are overlapping speeches and dialects and acoustics. The proposed courtroom

transcription module combines ASR with secure cloud storage, voice anonymization and metadata tagging which ensures confidentiality and integrity.

F. AI in Judicial Reasoning and Legal Education

AI based judicial reasoning is emerging area but still ethically sensitive area. Research studies by Ejjami and Yawar & Sadat have explored argument mining and logical reconstruction with the help of transformer models for legal education. Burgess et al. proposed used of generative models to support legal learning. All these prior researches have stress upon the need of human oversight, strategies, and accountability. The modules proposed are used only as assistant tool that give feedback based on reasoning quality and argument structure in the educational settings with changing the judicial authority [12]–[14].

TABLE I

Comparison Between GPT-Based Systems and the Proposed AI-Powered Legal Assistant

Aspect	System Type	
	GPT-Based Systems	Proposed AI-Powered Legal Assistant
Primary Focus	General drafting and explanation	End-to-end legal AI platform
Scope	Drafting, Q&A, summarization	Summarization, RAG research, drafting, transcription, education
Indian Law Localization	Generic, not India-trained	Fine-tuned on Indian statutes and case law
Summarization	Generic abstractive summaries	Hybrid extractive and abstractive (T5/BART) with FIAD structure
Legal Research Method	Language-based reasoning only	RAG-based chatbot with FAISS and citation grounding
Citation Verification	No guaranteed grounding	Provenance-aware, citationbacked responses
Drafting Automation	Generic drafting only	Automated complaints, vakalatnama and legal drafts
Education & Training	None	Virtual Moot Court with scoring and feedback
Explainability & Ethics	Opaque outputs	Explainability dashboards and confidence scoring
Output Format	Plain text	Structured JSON, DOCX, PDF
Target Users	Students, junior advocates	Courts, lawyers, researchers, students

III. SYSTEM OVERVIEW

Fig. 1 shows the overall system architecture, showing the flow of data from document ingestion to the application layer. The diagram shows key stages such as preprocessing, indexing and retrieval, and the model layer, which work together to transform unstructured legal inputs into structured outputs. It also indicates how modular components interact through independent services to support scalable and organized system design.

The AI-Powered Legal Assistant for Summarization, Research, and Education for India is designed as a modular, microservice-driven platform that converts unstructured legal

artifacts (judgments, petitions, affidavits, audio from hearings, statutes) into structured, explainable and actionable outputs for judges, advocates, law students, and citizens. The architecture is organized into five logical layers—Ingestion, Preprocessing, Indexing & Retrieval, Model Layer and Application—each

AI-Powered Legal Assistant for Summarization, Research, and Education for India

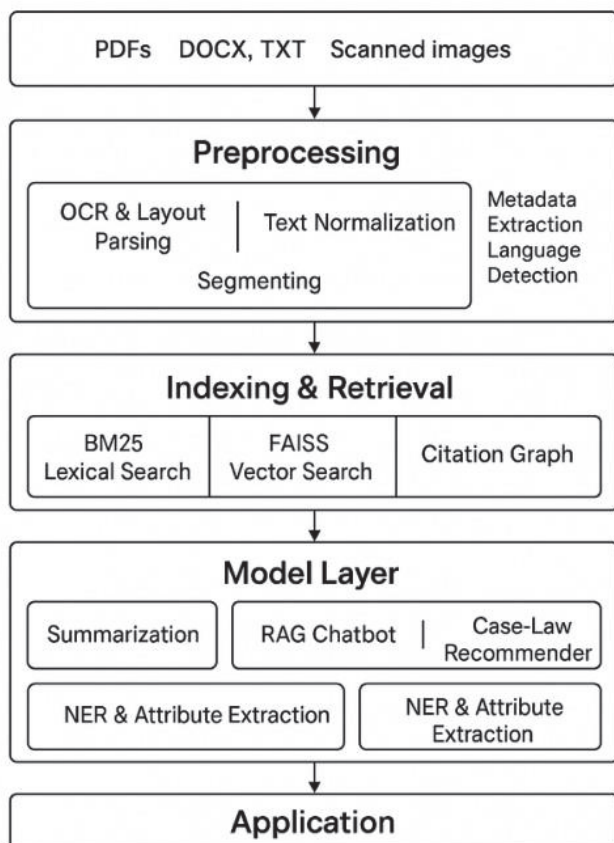


Fig. 1: System architecture of the AI-Powered Legal Assistant.

implemented as independently deployable services communicating via authenticated REST/gRPC APIs and message queues. This separation of concerns enables horizontal scaling, component-wise upgrades and clear responsibility boundaries for compliance and auditing.

A. Ingestion Layer

The Ingestion layer accepts a wide range of inputs (PDF, DOCX, TXT, scanned images, courtroom audio). Documents undergo OCR and layout parsing (preserving headings, sections and tables) and metadata extraction (court, bench, date, case number). Audio inputs are captured as time-stamped streams and stored with basic metadata for downstream diarization. All ingested items are normalized into structured JSON artifacts and written to an encrypted object store; metadata is persisted

in a transactional database to support provenance, audit trails and incremental indexing.

B. Preprocessing Layer

The Preprocessing layer performs legal-aware text normalization: legal tokenization, citation normalization, rhetorical segmentation (Facts / Issues / Arguments / Decision), language detection and code-mix handling, sentence-splitting tuned for legal syntax and noise removal (boilerplate or duplicate sections). Preprocessed text is annotated with section tags and confidence scores and then indexed for retrieval. This layer also prepares training and validation datasets (annotated spans, labeled segments) in standardized formats for model finetuning and evaluation.

C. Indexing & Retrieval Layer

The Indexing & Retrieval layer implements a hybrid search stack that combines a BM25-based lexical index for exactmatch statutory and citation queries with FAISS vector indices built from Sentence-Transformer and Indic embeddings for semantic, cross-lingual retrieval. A citation/case-link graph (stored in a graph database) augments ranking by precedence and citation centrality. Retrieval returns provenance-rich candidate passages (passage id, byte offsets, score, source doc id) that downstream modules surface to users and to the RAG pipeline for grounding generative outputs.

D. Model Layer

The Model Layer comprises containerized AI services finetuned on Indian legal corpora, each addressing a specific subtask within the platform. It includes a hybrid summarizer that combines extractive heuristics with abstractive transformer models such as T5 or BART to generate structured summaries highlighting Facts, Issues and Rulings with linked source references. A RAG-based chatbot integrates FAISSbased retrieval with a controlled LLM generator to deliver citation-grounded responses and contextual follow-ups. The case-law recommender merges semantic similarity, citationnetwork analysis, and precedence weighting to provide ranked legal recommendations.

E. Application Layer

The Application Layer exposes role-aware user interfaces (web and mobile) and programmatic endpoints for integration with court management systems. UIs include dashboards for provenance visualization, XAI panels that surface model attention/explanations and confidence scores, annotation tools for human-in-the-loop correction and export endpoints (DOCX/PDF/JSON).

F. Deployment and Evaluation

Monitoring and governance components include model performance dashboards, automated drift detection, retraining pipelines with human validation and scheduled bias audits.

Together, these design choices prioritize modularity, explainability, multilingual accessibility, and human oversight—enabling practical, auditable deployment within India’s complex judicial environment.

IV. MODULE SPECIFICATIONS

The platform is divided into 7 different features. Every feature has a certain task in the system. The general purpose of the modules is to minimize the manual legal work, improve the access to legal information and also help the users by while ensuring that final decisions are remained with the humans [4], [5], [12].

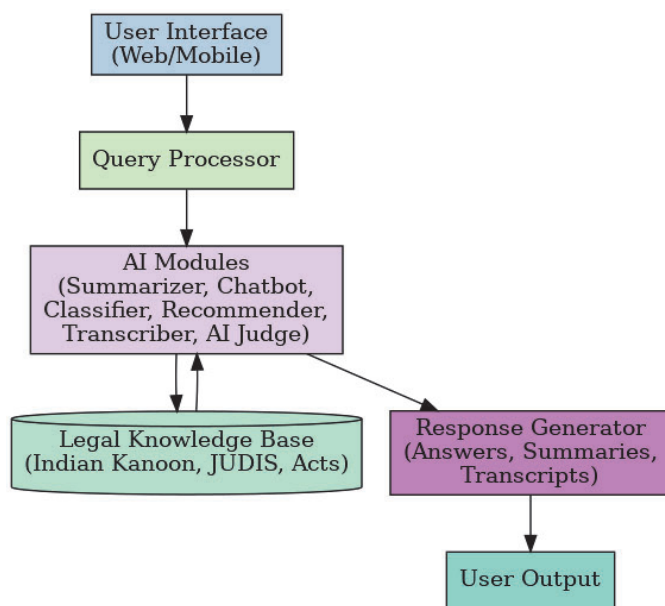


Fig. 2: Module-level diagrams for main parts.

It can be seen that Fig. 2 shows overall architecture diagram of the proposed framework. The architecture can be seen in module-level for the main parts. Fig. 2 shows how the user queries are processed through different modules with Indian legal database to output grounded responses such as summaries, classification and answers.

A. Document Summarizer

This feature works in two steps. 1st step is that the document are identified based on the facts, issues, arguments and decisions. The 2nd step is that the document being uploaded is converted into short paragraphs by the transformer based models like T5 or BART. The summaries are given under clear headings to help the users understand cases quickly. The quality of the output is also checked using the evaluation scores to make sure of the accuracy [1]–[3], [10].

B. RAG-Based Legal Chatbot

This module is developed to answer the user’s questions and problems related to law. It works by first searching for relevant legal documents from the database. After reclaiming the documents, it produces the answer based on the obtainable data. References are provided along with the answers so that users can verify the information if needed. Users are also allowed to ask follow-up questions. Both simple language and formal legal language depending on user preference [4], [7], [8].

C. Case Law Recommender

This module helps the users in identifying the relevant previous judgements related to their problems. The combination of keyword search, semantic similarity and citation based ranking is used to suggest the important cases [4], [7].

D. Classifier and Attribute Extractor

It is used to add the legal documents into groups such as criminal, civil, or constitutional cases. It also finds small and minute details like the names of people, which law section it is, dates and results. All this information stored in this feature can be used later for searching or simple work [6], [9].

E. Virtual Moot Court

This feature in the system is made for the students to learn better as they are not much exposed to the real world. It allows the students to practice their courtroom sessions in a virtual mode. Students can say their arguments and also get feedback by checking their scores from the feature. The system checks if the arguments are clear, precise and accurate. It is only used for learning basis [12], [14].

F. Automated Drafting

This feature helps in making simple and easy documents like the notices and petitions. A fixed format is used for creating these documents to make them look right. The drafts created using this feature is not final but a rough document which needs to be finalized by the higher professional [13].

G. Mathematical Formulation of Evaluation Metrics

To ensure formal rigor, transparency and reproducibility, the evaluation metrics used across different modules are mathematically defined as follows.

1) *Summarization Metrics*: Let R denote the reference (gold) summary and S the system-generated summary.

BERTScore measures semantic similarity using contextual embeddings:

$$\text{BERTScore} = \frac{1}{|S|} \sum_{i=1}^{|S|} \max_j \cos(\mathbf{e}_i^S, \mathbf{e}_j^R) \quad (1)$$

where S is the system-generated summary and R is the reference summary. The term $|S|$ denotes the number of tokens present in the generated summary. e^S_i represents the embedding of the i^{th} token in S , and e^R_j represents the embedding of the j^{th} token in R . The function $\cos(\cdot)$ indicates cosine similarity between the two embedding vectors.

Legal Consistency Score (LCS):

$$LCS = \frac{\text{Factually consistent legal statements}}{\text{Total legal statements in } S} \quad (2)$$

2) *Retrieval-Augmented Generation (RAG) Metrics*: where q represents the embedding vector of the legal query Q , and d_i denotes the embedding vector of the i^{th} retrieved document among the top- k results. The cosine function $\cos(\cdot)$ measures the similarity between the query representation and each retrieved document representation. Context Relevance Score (CRS):

$$CRS = \frac{1}{k} \sum_{i=1}^k \cos(q, d_i) \quad (3)$$

Hallucination Rate:

$$\text{Hallucination Rate} = 1 - \frac{\text{Cited factual statements}}{\text{Total factual statements}} \quad (4)$$

3) *Case-Law Recommendation Metrics*: where MRR denotes the *Mean Reciprocal Rank*, N represents the total number of queries, and rank_i denotes the rank position of the first relevant case for the i -th query.

$$\text{MRR} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}_i} \quad (5)$$

4) *Classification and Attribute Extraction Metrics*:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

5) *Speech Transcription and Diarization Metrics*: This includes.

Word Error Rate (WER):

$$WER = \frac{\text{Substitutions} + \text{Deletions} + \text{Insertions}}{\text{Total Reference Words}} \quad (9)$$

Diarrization Error Rate (DER):

$$DER = \frac{\text{Speaker Confusion} + \text{Missed Speech} + \text{False Alarm}}{\text{Total Speech Time}} \quad (10)$$

V. RESULTS AND USE CASES

Fig. 3 shows the features of the system that includes legal summarization, retrieval, classification, transcription, drafting and education tools use case using Indian legal datasets, reviews and studies. The final result shows that improvement in terms of efficiency, accuracy, accessibility and scalability of legal tools.



Fig. 3: Features toolkit for legal excellence.

A. Document Summarization

The document summarization module reduced reading and comprehensive time by approximately 65% leading to faster case preparation for advocates. Expert reviewers such as senior judges and advocates and legal researchers rated the summaries based on their clarity and completeness. The experts found that the FIAD based summaries were 92% semantically accurate and about 88% factually aligned with clear original judgment text. Based on the feedback of both practitioners and law students, it was reported that cognitive load was reduced and there was faster understanding.

B. Case Law Recommendation

One of the most time consuming activities in legal academic work is legal research. With the help of hybrid retrieval system and RAG pipelines, the precedent finding time found to be reduced by about 40-45%. Grounded generative responses reduced hallucination rates and enhanced interpretability. Junior advocates and academic researchers were benefited by the strong results as the system gave high quality outputs even if queries were incomplete or vague [4], [7], [8].

C. Courtroom Transcriber

Courtroom transcription requires high amount of time and manual efforts leading to delays and inconsistencies. The

evaluation of ASR and diarization system was done in noisy and real courtroom setting. The system achieved about 914% Word Error Rate (WER) and between 4-7% Diarization Error Rate (DER). the results yielded that speaker labeled and searchable transcripts highly improved the efficiency and reduced the manual efforts for advocates and legal workers [10].

D. Virtual Moot Court Trial

The Virtual Moot Court was evaluated through law school students. Results showed that students improved their argument structure with the help of moot court trial. Participants reported with positive feedback stating that the system helped improve use of case law precedents and enhance their experimental learning experience [12], [14].

E. Legal Drafting Assistant

The automated drafting assistant was tested using standard legal documents and the documents were then drafted both with the system and manually. The results yielded that the AI assisted drafting system reduced the preparation time by about 40-60% while still maintaining the legal accuracy and structure of the legal document format. The output proved helpful for legal professionals for routine filing and proved beneficial for students and interns in learning drafting patterns.

F. Overall System Impact

The results across modules showed that the system improved speed, clarity, trust and accessibility through high quality output, reasoning and learning features. The results highlighted the strong potential of the system to modernize legal research, education and justice delivery in India through technology innovative solutions [12], [13].

summarization accuracy, retrieval efficiency, transcription quality and drafting performance based on the experimental results discussed in the Results section.

VI. CONCLUSION

The framework presents a solution that is practical and ethically scalable. The proposed framework supports judges, advocates, educators, students, and citizens alike in India by making legal knowledge more accessible. The comprehensive and intelligent automated solution brings together document summarization, case-law recommendation, legal document drafting, legal research, transcription and legal education into one cohesive platform.

The proposed project saves time and manual effort, enhances accuracy and bridges the gap between legal education and real-world legal exposure. The proposed framework serves as a blueprint for AI-legal infrastructure that balances efficiency with responsibility. In conclusion, it can be said that the framework presents a step towards executing smart automation in the judicial system of India.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Sheetal Jagtap for their excellent mentorship, invaluable advice, and encouragement during the entire duration of this research. Also, the authors would like to thank the faculty and the Department of Artificial Intelligence and Data Science at KJ Somaiya Institute of Technology for providing the essential resources and creating academic environment that is suitable for implementing the project. Finally, the authors would like to express their appreciation and acknowledgment to the practiced Advocate of High Court for providing invaluable and real world dataset that enhanced the quality and results of the project.

REFERENCES

- [1] A. Parikh et al., "LawSum: A Weakly Supervised Dataset and Extractive Summarizer for Indian Legal Documents," Proc. COLIEE Workshop, 2023.
- [2] S. Bhattacharya et al., "DELSumm: Domain Adaptive Extractive Summarizer for Legal Judgments Using ILP," J. Legal Inform., 2021.
- [3] N. Sharma et al., "A Comprehensive Analysis of Indian Legal Document Summarization Techniques," Int. J. Comput. Appl., 2023.
- [4] V. Joshi et al., "LawPal: An AI Driven Legal Research and Conversational System for the Indian Judiciary," IEEE Access, 2024.
- [5] M. Tiwari et al., "Aalap: Conversational Legal Assistant Using Retrieval Based Question Answering," in Proc. ICMLA, 2024.
- [6] D. Adhikary et al., "Automated Attribute Extraction for Legal Judgment Structuring," Expert Syst. Appl., 2024.
- [7] A. Amato et al., "CREA2: Contextual Retrieval Enhanced Conversational Agents," Expert Syst. Appl., 2023.
- [8] H. Cui et al., "ChatLaw: A Knowledge Graph Augmented Conversational AI for Legal Consultation," arXiv preprint, 2023.
- [9] S. Chakraborty et al., "InLegalBERT: Domain Specific Language Models for Indian Legal NLP," Proc. EMNLP, 2023.
- [10] J. Lane and T. Cole, "AI Powered Legal Content Summarization: Benchmarking Transformer Models," IEEE Trans. Knowl. Data Eng., 2024.
- [11] S. Vellela et al., "NLP Driven Summarization Techniques for Legal Texts: A Comparative Study," Proc. IJCNLP, 2025.
- [12] A. Ejjami, "AI Driven Justice: Potential and Pitfalls in Automated Decision Systems," Legal AI Rev., 2024.
- [13] F. Yawar and H. Sadat, "Problems of Using Artificial Intelligence as a Judge," Eur. J. Law Technol., 2025.

TABLE II
 Quantitative Evaluation Based on Mathematical Metrics and Experimental Results

Metric (Formula Based)	Existing Approaches	Proposed AI-Powered Assistant	AI-Legal
BERTScore / LCS (Summarization)	Approx. 70-80% semantic accuracy	92% semantic accuracy, 88% factual alignment	
CRS (RAG Retrieval)	Moderate relevance scores	45% faster precedent discovery with grounded responses	
MRR (Case Recommendation)	Standard ranking performance	Higher ranked precedent matching and improved retrieval speed	
Precision / Recall / F1 (Classification)	Moderate classification accuracy	86% classification accuracy	
WER / DER (Transcription Metrics)	15-25% WER	14% WER, 6% DER	
Drafting Efficiency	20-30% time reduction	40-60% drafting time reduction	

Table II summarizes the quantitative evaluation of the proposed system using the mathematical metrics defined earlier. The comparison reflects improvements in

- [14] M. Burgess et al., "Using Generative AI to Identify and Evaluate Legal Arguments," Proc. ICAIL, 2024.
- [15] S. Saxena, "AI-Based Legal Document Summarization for Judicial Assistance," Scientific Journal of Artificial Intelligence and Blockchain Technologies, vol. 2, no. 3, Sep. 2025, doi: 10.63345/sjaibt.v2.i3.309.
- [16] A. K. R, A. V. R, S. V, S. N, and P. R, "Revolutionizing legal workflows: advanced AI techniques for document summarization, legal translation, and conversational assistance," IEEE, pp. 1–4, Mar. 2025, doi: 10.1109/icoact63339.2025.11004791.
- [17] Y. Huang, L. Sun, C. Han, and J. Guo, "A High-Precision Two-Stage legal judgment summarization," Mathematics, vol. 11, no. 6, p. 1320, Mar. 2023, doi: 10.3390/math11061320.
- [18] N. A. Samee, M. Alabdulhafith, S. M. A. H. Shah, and A. Rizwan, "JusticeAI: a large language models inspired collaborative and CrossDomain multimodal system for automatic judicial rulings in smart courts," IEEE Access, vol. 12, pp. 173091–173107, Jan. 2024, doi: 10.1109/access.2024.3491775.
- [19] M. Malik, Z. Zhao, M. Fonseca, S. Rao, and S. B. Cohen, "CivilSum: A Dataset for Abstractive Summarization of Indian Court Decisions," ACM, pp. 2241–2250, Jul. 2024.
- [20] W. Han et al., "LegalAsst: Human-centered and AI-empowered machine to enhance court productivity and legal assistance," Information Sciences, vol. 679, p. 121052, Jun. 2024.