iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77256

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Automated Legal Document Drafting in the Indian Judicial System: A Survey of AI-driven Approaches

Hemamalini S[1], Md Raqeeb Haider[2], Prajwal Jahagirdar[3], Shashwat Kumar[4], Dr. Rohini T V[5]
Computer Science & Engineering Dayananda Sagar College of Engineering Bengaluru, India

Abstract: ***The Indian judicial system faces many challenges which slow down legal document preparation because of the complex procedures that is present in judicial system, numerous cases and language differences. The process of manual transcription of different languages, extraction of entities and court standard compliance formatting needs help of lawyers to spend a lot of time which results in late and incorrect work. The research states that we have done evaluation of 20 academic papers from 2017 to 2025 which aims to investigate Automatic Speech Recognition (ASR), Natural Language Processing (NLP) and transfer learning models for legal systems. The research shows that Whisper ASR models achieve a high level performance with help of 55.2% error reduction from previous models but these improvements focus on Western regions that use languages with a lot of resources. Our current state-of-the-art models show very low applicability for Indian legal systems because they were developed for Western jurisdictions with high-resource languages. The current systems does not have essential features because they do not handle issues like overlapping speech and they do not integrate multiple techniques like multilingual transcription, speaker diarization, legal entity extraction and jurisdiction- aware template generation. The Indian judicial system requires solutions to these problems because they create problems for the public access to justice and prevent judicial advancements.***
Index Terms: Legal NLP, Speech Recognition, Multilingual Systems, Indian Judiciary, Survey, AI Automation.

I. INTRODUCTION

Every judicial system across the world needs legal doc- ument preparation as an operational necessity. The Indian legal system of India faces many obstacles because it operates under multiple jurisdictions with separate state procedures and official languages and dialects. The present manual system for note-taking and spoken transcription and cross-verification produces human mistakes which lead to operational delays and result in millions of unresolved cases. The implementation of automated systems which convert spoken dialogue into court documents will reduce judicial workloads and make court documents more accessible to the public.

The integration of Automatic Speech Recognition (ASR) with Natural Language Processing (NLP) technologies pro- duces substantial transformations in judicial operational sys- tems. The current tools demonstrate three major limitations because they fail to recognize legal terminology and distin- guish between different speakers and extract relevant legal information from noisy multilingual data. The development of customized integrated solutions stands as the important requirement for this purpose.

A. Scope and Contributions

The research evaluates present technological progress for building intelligent systems which generate legal documents automatically. The research examines four important elements which include multilingual Automatic Speech Recognition systems, speaker identification, legal domain-specific Natural Language Processing and parameter-efficient model optimiza- tion.

The survey employs a systematic approach to evaluate automated legal document drafting systems while identifying their existing knowledge deficiencies. The first part of this paper describes the problems with manual legal document preparation in India's multilingual judicial system and defines the research parameters.

The review section divides studied papers into five core technical elements which include ASR and Diarization and Domain-Specific Legal NLP and Transfer Learning and Dataset Creation and Resource Scarcity and Model Efficiency and Integrated Systems. The current systems face major obstacles which prevent them from scaling up their solutions for the Indian market. The study provides critical results about integrated systems before describing upcoming research that will expand datasets and assess ethical systems and create new workflows.

The survey has the following contributions:

1) The research studies consist of four essential technical blocks which serve as the base for building complete legal automation systems.
2) The assessment of performance results shows particular research gaps because the Indian legal system operates with multiple languages and multiple court systems
3) The Summary of Findings and Research Gaps section de- tails existing work restrictions because of limited datasets and absent integrated systems.
4) The research investigates LegalTech development require- ments by studying how to adapt across languages and build larger datasets and deploy ethical AI systems.

II. RELATED WORK

Engineering automated legal document drafting platforms involves multiple technologies requiring integration. This sec- tion organizes essential papers into technical components of proposed systems.

A. ASR and Diarization for Judicial Systems

Automation processes start with spoken audio conversion into structured transcripts including speaker identification in- formation.

1) Robust Multilingual ASR: Development of large-scale pre-training methods brought major improvements in ASR technology. Whisper system (Radford et al., 2022) [1] repre- sents important breakthroughs using 680,000 hours of weakly supervised web audio developing a multilingual encoder- decoder transformer. The system achieved 55.2% error rate reduction compared to wav2vec 2.0 when tested on multiple datasets. The system shows built-in resistance to environmen- tal noise and speaker variation suitable for legal settings. The model requires legal terminology adjustments lacking domain knowledge. FLEURS benchmark (Conneau et al., 2022) [3] reveals performance differences between languages with limited resources including Indian courtroom languages.
2) Speaker Diarization: Speaker diarization needs exact methods achieving judicial approval and maintaining docu- ment consistency. Wang et al. (2017) [5] showed deep embed- dings (d-vectors) work best with clustering methods achieving strong speaker identification. Lyra et al. (2024) [2] developed integrated systems combining Kaldi ASR toolkit with d-vector diarization processing Brazilian court recordings. Research showed legal terminology and speech overlap presented major obstacles. Modern toolkits PyAnnote and NVIDIA NeMo support advanced deep embedding techniques.

B. Domain-Specific Legal NLP and Transfer Learning

Recorded audio needs conversion into organized data fol- lowing legal standards.

1) Legal Entity Recognition and Classification: General- purpose NLP models show weak performance handling spe- cialized legal operations. Chalkidis et al. (2020) [4] de- veloped Legal-BERT using large English legal corpus for pre-training. Legal-BERT demonstrates superior performance than general BERT models performing legal tasks including classification and entity recognition. Kalamkar et al. (2022) [23] built specialized NER system for Indian court judgment processing detecting 14 legal entities achieving high F1 scores. IndicBERT model from Kakwani et al. (2020) [7] established solid base for Indian language processing yet requires legal domain training.
2) Document Structuring and Reasoning: Legal document understanding depends on document organization and logical reasoning abilities. Indian Legal Information Retrieval System (ILIRS) uses property graphs linking semantic elements per- forming better than BERT/BiLSTM models in entity recog- nition tasks. Legal AI systems need structured data models enabling them to surpass current text classification capabilities.

C. Dataset Creation and Resource Scarcity

Legal research faces challenges due to insufficient high- quality annotated legal datasets. Indian judiciary faces extreme shortage of available open legal datasets.

1) Dataset Benchmarking: Joshi et al. (2024) [11] created IL-TUR as first complete benchmark for Indian legal NLP. Research indicates SOTA models including GPT-4 achieve only average results reaching 66.67% mF1 accuracy on BAIL task.
2) Dataset Generation: Deshmukh and Kamble (2025) [12] developed IndianBailJudgments-1200 through LLM pipeline annotation of 1,200 bail orders proving efficient method for large-scale dataset creation.

D. Model Efficiency and Integrated Systems

Deployment of practical scalable solutions for district courts depends on model efficiency as essential foundation.

1) Parameter-Efficient Fine-Tuning (PEFT): Hu et al. (2022) [8] presented LoRA (Low-Rank Adaptation) decreasing trainable parameters through weight freezing and small train- able rank-decomposition matrices injection. Method allows Whisper model adaptation for Indian languages and legal terminology needing minimal computational power.
2) Integrated Solutions: Rahman et al. (2024) [16] and Ariai et al. (2025) [27] established contract analysis as leading legal NLP application yet researchers need advancing litigation support and complete court transcription systems. Research addresses this knowledge deficiency requiring integrated sys- tems uniting modular components (ASR, Diarization, NLP, Template Generation).

III. PROBLEM STATEMENT

The legal sector requires extensive documentation which leads to repetitive information transcription and organization of client–lawyer dialogues and witness statements and hearing recordings. The present work environment requires extended periods of operation which produces multiple errors that re- quire extensive human intervention for correction. The current transcription tools lack ability to process multiple languages and they fail to recognize speech that overlaps or contains background noise. The extraction of essential legal data needs advanced evaluation methods which prevent automated sys- tems from being implemented. The current systems fail to convert spoken dialogue into legal documents which meet all necessary jurisdictional standards.

The project implements an AI-based automation system which uses speech recognition and speaker diarization and natural language processing and document parsing to generate legal documents at higher speeds.

IV. LITERATURE REVIEW

A. Speech Recognition and Diarization

1) Robust Speech Recognition via Large-Scale Weak Su- pervision - A. Radford et al.: The seq2seq Transformer model known as Whisper received training from 680,000 hours of web audio data for ASR and translation and language ID tasks but it shows strong noise resistance yet lacks legal terminology and speaker identification capabilities.
2) Automatic Transcription of Court Hearings: A Game Changer for Court Hearings. - A. Lyra et al.: The Kaldi ASR system with d-vector diarization technology successfully detected roles in Brazilian courtroom record- ings but it struggled with speech overlap and language domain adaptation issues.
3) FLEURS: Few-Shot Learning Evaluation of Universal Representations of Speech - A. Conneau et al.: The 102-language benchmark evaluates few-shot ASR perfor- mance which demonstrates that models achieve lower results when working with languages that have limited available data.
4) Speaker Diarization with Deep neural network em- beddings - D. Garcia-Romero et al.: The system uses d-vectors to extract speaker information which produces effective results but the method becomes less accurate when dealing with multiple speakers who speak at the same time. The research investigates essential methods which speech processing requires. The Whisper model Whisper from Rad- ford et al. The system operates as a complete ASR system ac- cording to Lyra et al. The system performs d-vector diarization for courtroom analysis through its implementation of Kaldi. The research by Garcia-Romero et al. demonstrates d-vector- based speaker diarization methods. The FLEURS dataset from Conneau et al. The system enables researchers to evaluate their models at different language resource levels.
5) Similarities: The system transforms audio data into orga- nized output formats through diarization which operates using d-vector embeddings for its processing. The system generates substandard results when it encounters technical terms and operates with limited language resources.
6) Differences: The Whisper system functions as a wide general framework yet Lyra implements Kaldi for particular domain operations and Garcia-Romero works on diarization tasks and FLEURS serves as a reference dataset.
7) Limitations: The system faces three main challenges which include its inability to recognize legal terminology and its weak performance with multilingual Indian languages and its failure to separate speakers when their voices overlap.

B. Legal-Specific NLP Models and Benchmarks

1) Legal-BERT: The Muppets straight out of Law School
2) I. Chalkidis et al.: The BERT model receives pre- training from EU legal texts which enhances its perfor- mance on EU legal tasks yet it does not work for Indian or multilingual legal domains.

3) Neural Legal Judgment Prediction in English - I. Chalkidis et al.: The system uses transformer models to predict legal outcomes and statutes from well-organized Western case data but it requires structured text input.
4) Indic-transformers: An analysis of transformer lan- guage models for indian languages - K. Jain et al.: The research team developed BERT-style models which achieved better results for Indian language processing yet they did not receive legal domain-specific training.
5) IL-TUR: Benchmark for Indian Legal Text Under- standing and Reasoning - A. Joshi et al.: The English- Hindi benchmark evaluates LLM performance through expert-annotated data which produces standard output that needs advanced models to produce superior results.
6) IndianBailJudgements-1200: A Multi-Attribute Dataset for Legal NLP on Indian Bail Order - S. Deshmukh et al.:The multi-attribute bail order dataset enables researchers to predict bail order attributes through a system which maintains particular task requirements.
7) Unlocking the Future: The Intersection of Artificial Intelligence and the Juducial System - R. Barman et al.: The system performs rhetorical-role segmentation on 50,000 Indian decisions which helps with preprocessing but needs additional integration and annotation work.

The research group transitions from basic NLP to legal domain models and datasets to prove the requirement for domain-specific and multilingual resources.

a) Similarities: The performance of models improves when they receive domain-specific pre-training or access to specific datasets and most models employ Transformer architectures and demonstrate the need for more Indian legal data in multiple languages.
b) Differences: The research includes two types of pre-training models (Legal-BERT and Indic-transformers) and two types of benchmark and dataset resources (IL-TUR and Bail dataset); the research focuses on different legal jurisdictions.
c) Limitations: The models trained in Western countries fail to achieve good results in Indian settings because India lacks sufficient diverse legal resources and high annotation expenses prevent large-scale deployment.

C. *General Language Model Adaptation*

1) Low-Rank Adaptation of Large Language Models (LoRA) - E. J. Hu et al.: The method enables efficient fine-tuning through the addition of low-rank adapters which protect the base weights from changes.
2) A Simple and Effective Method to Improve Zero-Shot Cross-Lingual Transfer Learning - K. Ding et al.: The CLLP method enables zero-shot transfer improvement through English space mapping of low-resource language representations without requiring additional data.
3) Exploring the Limits of Transfer Learning with a Uni- fied Text-to-Text Transformer (T5) - C. Raffel et al.: The text-to-text framework T5 demonstrates exceptional performance on text tasks yet it lacks the ability to handle speech data.

The section examines two vital approaches for parameter- efficient and cross-lingual adaptation which allow big models to function in restricted domain-specific settings.

Similarities: All methods provide solutions to enhance model adaptability and operational efficiency.

Differences: The three approaches operate through distinct methods because LoRA enhances model fine-tuning speed and CLLP enhances language transfer abilities and T5 operates independently for text processing.

Limitations: The method needs domain-specific data to operate LoRA but CLLP shows inferior results to in-language fine-tuning and T5 lacks the ability to process speech.

D. *Legal NLP Applications and Techniques*

1) Analysing similarities between legal court documents using natural language processing approaches based on transformers - R. S. D. Oliveira et al.: The method uses transformer embeddings together with clustering for large document similarity analysis but requires substantial resources and remains limited to specific languages.
2) Natural Language Understanding (NLU)-Powered In- formation Retrieval from Indian Legal Property Graph - K. Shreeram et al.: The system constructs a legal property graph which enables NLU queries to exe- cute as graph queries for enhanced retrieval performance but depends on precise extraction results.
3) Natural language processing in legal document anal- ysis software: A systematic review of current ap- proaches, challenges, and opportunities - M. M. Rah- man et al.: The authors evaluate existing methods which demonstrate transformer superiority but they point out two major limitations: insufficient data availability and inadequate evaluation criteria.
4) Enforcing legal information extraction through context-aware techniques: The ASKE approach - S. Castano et al.: The method

uses zero-shot learning with context-aware embeddings to perform iterative extraction which produces better results although it faces challenges with ambiguous information.

5) Leda: A System for Legal Data Annotation - S. Ad- hikary et al.: The web-based annotation system allows users to handle changing legal concepts and agreement metrics through its quick annotation process which needs multilingual evaluation.

6) The Impact of Large Language Modeling on Natural Language Processing in Legal Texts: A Comprehen- sive Survey - D. H. Anh et al.: The research investigates how LLMs affect legal text processing which provides benefits yet encounters difficulties in understanding in- terpretations and detecting biased content.

7) LetSum, an Automatic Legal Text Summarizing -

a) Farzindar et al.: The system produces standardized summaries from legal documents through automated pro- cessing but produces variable results when processing texts from different domains and language backgrounds. The group shows functional systems and research evidence that proves transformer advancements yet they encounter difficulties with data management and term processing and language support for multiple languages.

b) Similarities: The systems operate through transformer tech- nology to automate legal work yet they face challenges be- cause of limited labeled data and restricted domain terminol- ogy.

c) Differences: The system includes multiple modules which perform clustering and graph retrieval and annotation tools and summarization functions while operating across different jurisdictions and languages.

d) Limitations: The systems function under particular lan- guage rules and legal systems but extraction errors create oper- ational problems because researchers need more multilingual annotated data for complete analysis.

TABLE I
Overview of Surveys in NLP in the Legal Domain

References	Covered Subjects	Year	Domain Focus	Low-Res. Lang.	Integrated Systems
Rahman et al. [16]	✓	2024	✗	✗	✗
Anh et al. [19]	✓	2023	✗	✗	✗
Krasadakis et al. [21]	✓	2024	✓	✓	✗
Ariai et al. [27]	✓	2025	✓	✗	✗
Mahari et al. [28]	✓	2023	✓	✗	✗
Ong et al. [36]	✓	2024	✓	✓	✗

TABLE II
Overview of Indian Legal Datasets and Benchmarks

Dataset/ Benchmark	Description	Lang.	Primary Task(s)	Ref.
IL-TUR	First comprehensive benchmark for In- dian legal NLP (8 tasks)	Eng., Hindi	Multi-Task (Sum., QA, NER)	Joshi et al. [11]
IndianBail Judgments- 1200	1,200 annotated bail orders	English	Text Class., Le- gal Reasoning	Deshmukh & Kamble [7]
Multi-Attr. Bail Dataset	3,500+ bail orders with 6 attributes	English	Multi-Label Class.	Poddar et al. [12]
LegalSeg	50,000+ judgments with rhetorical roles	English	Doc. Segmenta- tion	Maheshwari et al. [13]
MILDSum	3,122 judgments with Eng/Hi summaries	Eng., Hindi	Summarization	Datta et al. [32]
Indian Legal NER Corpus	46,545 legal entities (14 types)	English	NER	Kalamkar et al. [23]
Indian Legal DB (ILDC)	Judgments with ex- pert summaries	English	Sum., Simplifi- cation	Swamy et al. [33]

V. LIMITATIONS AND RESEARCH GAPS

Current state of individual NLP and ASR components shows significant advancement yet researchers have not developed working solutions for multilingual Indian judicial systems. Reviewed literature reveals multiple research gaps and system limitations.

A. Key Research Gaps

1) ASR for Legal Domain: The research by (Radford et al., 2022) [1] provides outstanding multilingual transcription abilities which lead to a 55.2% performance improvement above the current best average. The system encounters two pri- mary difficulties because it fails to identify legal terminology, complex vocabulary and because it lacks ability to handle si- multaneous dialogues between multiple speakers. The training data does not have sufficient Indian language representation because these languages have restricted availability. The up- coming versions of Whisper require LoRA adapter integration to achieve better performance in particular application domains through language translation functionality.

2) Speaker Diarization: The D-vector and neural methods from (Wang et al., 2017) [5] serve as fundamental components for PyAnnote/NVIDIA NeMo tools in speaker separation technology. The system performance deteriorates to an ex- treme level when it encounters background noise and simul- taneous speech which occurs during real court proceedings. The systems need PyAnnote and NVIDIA NeMo advanced diarization tools to detect speakers when audio recordings contain background noise.

3) Legal NLP & Structuring: Legal-BERT models outper- form general models when performing legal classification tasks because they achieve superior results. The analysis of Indian court decisions through structural methods reaches an accuracy rate of 84%. The present solutions use text-based interfaces which do not identify legal boundaries that affect automated document generation. The systems start performing analytical tasks instead of producing complete documents. AI systems of the future require agents to access data first before they verify results and produce output through a template engine which operates under jurisdictional rules.

4) Datasets & Resources: The LLM-based workflow produces structured data collections at high speed for the IndianBailJudgments-1200 dataset. The lack of open legal datasets with annotations makes it impossible for researchers to access Indian legal cases which exist beyond English and Hindi language boundaries. The development of new datasets requires two main priorities which include adding Indian case examples and implementing multiple language capabilities.

5) Integrated Systems: The research studies each individ- ual component (ASR, NER, Summarization) independently. The reviewed literature lacks any paper that demonstrates a complete system which converts spoken audio into properly formatted legal documents following Indian jurisdiction rules of jurisdiction. The solution requires developers to build an integrated system which connects all necessary components.

TABLE III
Research Gap Analysis

Aspect	Existing Gap	Required Solutions
Multi-lingual	33.3% surveys + 28.6% datasets cover low-res. langs	Domain-specific fine-tuning for Indian languages
Speech	No ASR coverage in sur- veys/datasets	Integrated ASR with legal terminology
End-to-End	0% integrated systems	Complete audio-to-doc pipeline
Jurisdiction	Generic datasets, limited fo- cus	State-specific template systems
Automation	Partial coverage, analysis only	Complete document generation
Domain	66.7% domain, task-specific	Comprehensive legal fine-tuning
Key Finding: Existing research lacks integrated, multilingual, speech-enabled sys- tems for legal automation. Future work must address end-to-end pipeline development combining ASR, NLP, and jurisdiction-aware systems.		

VI. CHALLENGES AND OPPORTUNITIES

Even with proven research outcomes, there remain several obstacles and opportunities for further study and advancement:

1) Multilingual Complexity and Dialect Adaptation: The process to achieve equal transcription accuracy for all Indian languages continues to be difficult to achieve. The models need ongoing updates to handle different dialects and accent patterns and code-switching between languages. Research activities should concentrate on de- veloping training materials that align with dialects and enhancing language adaptation systems for minority lan- guages.

2) Bias and Fairness: The development process for le- gal entity extraction and document analysis algorithms requires absolute impartiality. The process of incorpo- rating multiple legal frameworks and minority groups and local court procedures proves difficult because it requires avoiding any form of discrimination. Research efforts should focus on creating ethical auditing methods and training systems which teach professionals to uphold fairness in their professional activities.

3) Dataset Scarcity and Annotation Quality: The col- lection of high-quality annotated legal data for Indian court cases faces a significant challenge because bail judgment data represents the only available resource. The creation of large legal datasets which receive professional legal verification represents an essential requirement. The combination of semi-supervised learning methods with synthetic data creation techniques helps solve dataset shortages while preserving high-quality annotations.

4) Scalability and Integration with E-Court Systems: The system needs to optimize its speech processing operations and storage systems and real-time transcription processing to handle large court proceedings effectively. Research needs to address policy and technical obstacles which block the integration of government e-filing portals with judicial infrastructure systems through interoperabil- ity.

5) Explainability and Legal Accountability: The system needs to display its operations clearly while humans must review outputs to verify results for building trust in AI-based legal systems. The legal profession requires research-based explainable AI solutions to verify and understand automated legal recommendation systems.

VII. CONCLUSION

The survey provides a complete assessment of AI-based systems which generate legal documents automatically for the Indian judicial system. The evaluation of 36 papers spanning from 2017 to 2025 shows major progress in ASR technology because Whisper achieved a 55.2% error decrease and Legal- BERT delivered improved results for legal NLP applications in specific domains. The current technology fails to provide solutions that meet the particular needs of the Indian judicial system.

The research shows that Indian language datasets have insufficient data and speaker diarization systems fail to work properly with noisy audio because of legal restrictions against complete end-to-end automation. The research identifies three main gaps in current studies. The surveys and datasets fail to support enough Indian languages because they only cover 33.3% of surveys and 28.6% datasets address low- resource languages. The reviewed work lacks complete audio- to-document pipelines because no study presents integrated systems. The current systems fail to adapt their legal terminol- ogy and procedures to the specific requirements of the Indian legal domain. The research must fulfill three essential goals which in- clude (1) creating professionally identified datasets for In- dian legal procedures and languages through IL-TUR and IndianBailJudgments-1200 models; (2) The system will un- dergo ethical and fairness assessments through state-of-the-art bias detection and model interpretability tools which preserve legal responsibility; (3) The system needs workflow integration to perform automated e-filing through direct access to the E- Court System; (4) The system needs scalability improvement through lightweight fine-tuning methods which work well for courts with restricted funding. The development of integrated systems which merge speech recognition with NLP and jurisdiction-specific components will establish efficient legal documentation solutions that work for all linguistic groups in India.

REFERENCES

[1] Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey and I. Sutskever, ”Robust speech recognition via large-scale weak supervision,” in Proc. Int. Conf. on Machine Learning (ICML), pp. 28492–28518, PMLR, July 2023.

[2] A. Lyra, C. E. Barbosa, H. Salazar, M. Argoˆlo, Y. Lima, R. Motta and J. M. de Souza, ”Automatic Transcription Systems: A Game Changer for Court Hearings,” 2024.

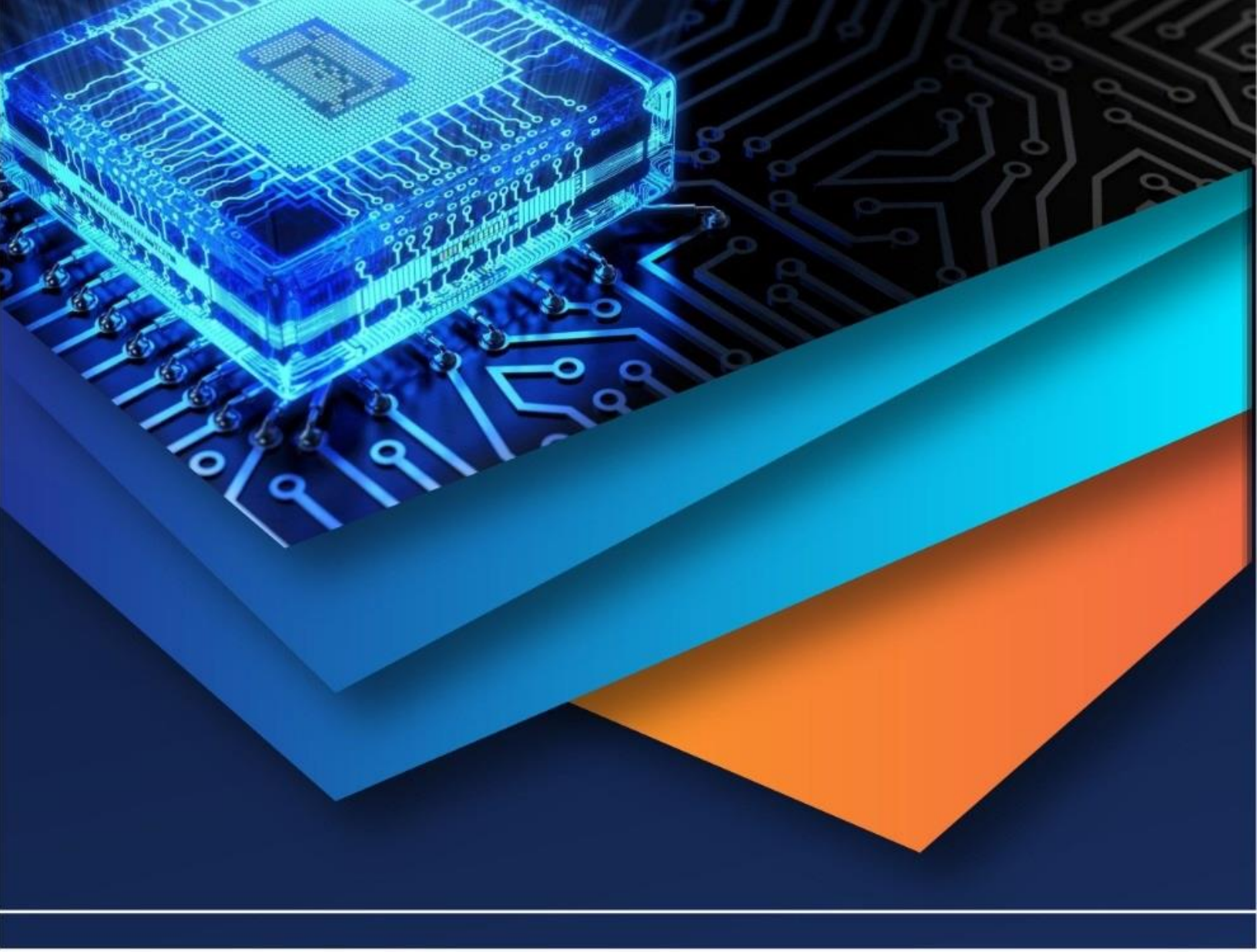
[3] A. Conneau, et al., ”Fleurs: Few-shot learning evaluation of universal representations of speech,” in 2022 IEEE Spoken Language Technology Workshop (SLT), pp. 798–805, IEEE, January 2023.

[4] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras and I. Androut- sopoulos, ”LEGAL-BERT: The muppets straight out of law school,” arXiv preprint arXiv:2010.02559, 2020.

[5] D. Garcia-Romero, D. Snyder, G. Sell, D. Povey and A. McCree, ”Speaker diarization using deep neural network embeddings,” in 2017 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 4930–4934, IEEE, March 2017.

[6] I. Chalkidis, I. Androutsopoulos and N. Aletras, ”Neural legal judgment prediction in English,” arXiv preprint arXiv:1906.02059, 2019.

[7] K. Jain, A. Deshpande, K. Shridhar, F. Laumann and A. Dash, "Indic- transformers: An analysis of transformer language models for indian languages," arXiv preprint arXiv:2011.02323, 2020.
[8] E. J. Hu, et al., "Lora: Low-rank adaptation of large language models," in ICLR, vol. 1, no. 2, p. 3, 2022.
[9] K. Ding, et al., "A simple and effective method to improve zero-shot cross-lingual transfer learning," arXiv preprint arXiv:2210.09934, 2022.
[10] C. Raffel, et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," in Journal of Machine Learning Research, vol. 21, no. 140, pp. 1–67, 2020.
[11] A. Joshi, S. Paul, A. Sharma, P. Goyal, S. Ghosh and A. Modi, "IL- TUR: Benchmark for Indian legal text understanding and reasoning," arXiv preprint arXiv:2407.05399, 2024.
[12] S. Deshmukh and P. Kamble, "IndianBailJudgments-1200: A Multi- Attribute Dataset for Legal NLP on Indian Bail Orders," arXiv preprint arXiv:2507.02506, 2025.
[13] R. Barman, "Unveiling the Future: The Intersection of Artificial Intelli- gence and the Judicial System," in Indian J. Integrated Rsch. L., vol. 3, p. 1, 2023.
[14] R. S. D. Oliveira and E. G. Sperandio Nascimento, "Analysing similar- ities between legal court documents using natural language processing approaches based on transformers," in PloS One, vol. 20, no. 4, e0320244, 2025.
[15] K. Shreeram, A. Patil, A. Agarwal, D. P. Rana and R. G. Mehta, "Natural Language Understanding (NLU)-Powered Information Retrieval from Indian Legal Property Graph," in Procedia Computer Science, vol. 258, pp. 2500–2509, 2025.
[16] M. M. Rahman, N. M. Gony, M. M. Rahman, M. Rahman and S. D. Khatun Shuvra, "Natural language processing in legal document analysis software: A systematic review of current approaches, challenges, and opportunities," in Int. Journal of Innovative Research and Scientific Studies, vol. 8, no. 3, pp. 5026–5042, 2025.
[17] S. Castano, et al., "Enforcing legal information extraction through context-aware techniques: The ASKE approach," in Computer Law & Security Review, vol. 52, 105903, 2024.
[18] S. Adhikary, D. Roy, D. Ganguly, S. Kumar Guha and K. Ghosh, "Leda: a system for legal data annotation," in Frontiers in Artificial Intelligence and Applications, no. 379, pp. 370–367, 2023.
[19] D. H. Anh, D. T. Do, V. Tran and N. Le Minh, "The impact of large language modeling on natural language processing in legal texts: A comprehensive survey," in 2023 15th Int. Conf. on Knowledge and Systems Engineering (KSE), pp. 1–7, IEEE, October 2023.
[20] A. Farzindar and G. Lapalme, "Letsum, an automatic legal text summa- rizing," in Legal knowledge and information systems: JURIX, vol. 11, 2004.
[21] P. Krasadakis, E. Sakkopoulos and V. S. Verykios, "A survey on challenges and advances in natural language processing with a focus on legal informatics and low-resource languages," in Electronics, vol. 13, no. 3, p. 648, 2024.
[22] M. Yang, et al., "Low-rank adaptation for foundation models: A com- prehensive review," arXiv preprint arXiv:2501.00365, 2024.
[23] P. Kalamkar, A. Agarwal, A. Tiwari, S. Gupta, S. Karn and V. Raghavan, "Named entity recognition in Indian court judgments," arXiv preprint arXiv:2211.03442, 2022.
[24] I. Chalkidis, M. Fergadiotis and I. Androutsopoulos, "MultiEURLEX– A multi-lingual and multi-label legal document classification dataset for zero-shot cross-lingual transfer," arXiv preprint arXiv:2109.00904, 2021.
[25] R. M. R. Kadiyala, S. Pullakhandam, K. Mehreen, S. Tippareddy and A. Srivastava, "Augmenting Legal Decision Support Systems with LLM-based NLI for Analyzing Social Media Evidence," arXiv preprint arXiv:2410.15990, 2024.
[26] T. Y. S. S. Santosh, K. D. Ashley, K. Atkinson and M. Grabmair, "Towards supporting legal argumentation with NLP: Is more data really all you need?," in Proc. of the Natural Legal Language Processing Workshop 2024, pp. 404–421, November 2024.
[27] F. Ariai, J. Mackenzie and G. Demartini, "Natural language processing for the legal domain: A survey of tasks, datasets, models, and chal- lenges," arXiv preprint arXiv:2410.21306, 2024.
[28] R. Mahari, D. Stammbach, E. Ash and A. Pentland, "The law and NLP: Bridging disciplinary disconnects," in Findings of the Association for Computational Linguistics: EMNLP 2023, pp. 3445–3454, December 2023.
[29] R. Christen, A. Shaitarova, M. Stu¨rmer and J. Niklaus, "Resolving legalese: A multilingual exploration of negation scope resolution in legal documents," in Proc. of the 2024 Joint Int. Conf. on Computational Lin- guistics, Language Resources and Evaluation (LREC-COLING 2024), pp. 13992–14004, May 2024.
[30] A. Pahilajani, S. Jain and D. Trivedi, "NLP at UC Santa Cruz at SemEval-2024 task 5: Legal answer validation using few-shot multi- choice QA," in Proc. of the 18th Int. Workshop on Semantic Evaluation (SemEval-2024), pp. 1309–1314, June 2024.
[31] J. Li, R. Bhambhoria, S. Dahan and X. Zhu, "Experimenting with legal ai solutions: The case of question-answering for access to justice," arXiv preprint arXiv:2409.07713, 2024.
[32] D. Datta, S. Soni, R. Mukherjee and S. Ghosh, "MILDSum: A novel benchmark dataset for multilingual summarization of Indian legal case judgments," arXiv preprint arXiv:2310.18600, 2023.
[33] K. Swamy, O. Salgare, O. Sakhare and S. Tamboli, "ValidEase: NLP for Simplification and Summarization of Legal Documents," 2024.
[34] J. J. W. Ng, et al., "Evaluating the performance of artificial intelligence- based speech recognition for clinical documentation: a systematic re- view," in BMC Medical Informatics and Decision Making, vol. 25, no. 1, p. 236, 2025.
[35] M. M. Rahman, N. Gony, M. M. Rahman, M. M. Rahman and M. SD, "Natural language processing in legal document analysis: A systematic review of current approaches," in Int. J. Innov. Res. Sci. Stud., vol. 8, pp. 5016–5072, 2025.
[36] D. M. Katz, D. Hartung, L. Gerlach, A. Jana and M. J. Bommarito II, "Natural language processing in the legal domain," arXiv preprint arXiv:2302.12039, 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY