

ADVANCED NATURAL LANGUAGE PROCESSING FOR LEGAL DOCUMENT  
ANALYSIS

Srinivasa Kalyan Vangibhurathachhi  
Srinivasa.Kalyan2627@gmail.com

---

*Abstract*

*The exponential growth of complex legal documents that includes statutes, contracts and case law has created significant bottlenecks for legal professionals relying on manual review and traditional summarisation techniques. Traditional methods such as keyword extraction, sentence reduction and rule-based summarisation fall short in capturing the rich semantics and context of legal texts resulting in inefficiencies and human error. This research critically examines the integration of advanced Natural Language Processing (NLP) techniques in legal document analysis as a transformative solution to these challenges.*

*Key research findings indicated that NLP models such as Legal-BERT, LegalPro-BERT and RoBERTa, when fine-tuned on legal corpora, significantly outperform general-purpose models in tasks such as entity recognition, clause extraction, and multi-label classification. The research study also finds that domain-specific summarisation tools, while improving content condensation, still struggle with accuracy in summarising complex legal arguments. NLP applications in contract review, legal research, AI-powered chatbots and predictive analytics have demonstrably reduced manual workloads and improved decision-making efficiency. However, the research identifies persistent challenges including data privacy concerns, algorithmic bias from historical training data, and the lack of high-quality annotated legal datasets. Future trends suggest shift towards self-learning, multilingual, and real-time legal NLP systems. In summary, while NLP holds immense potential to enhance efficiency and access to justice, its deployment requires robust ethical frameworks and interdisciplinary collaboration to ensure legal validity, transparency, and accountability.*

*Keywords: Natural Language Processing, Legal Document Analysis, Named Entity Recognition; NER, text classification, legal domain, judgement prediction, contract review, legal research, predictive legal analytics*

## I. INTRODUCTION

The legal profession is often overburdened by large volumes of complex textual data ranging from statutes, case laws to contracts, regulations and legal opinions. With the exponential growth of legal documents and the urgent need to effectively extract information, there has

been a growing challenge in summarizing legal texts for contemporary legal practice and research (Jagirdar et al., 2024). Accordingly, the increasing volumes and complexity of legal documents tend to overwhelm legal experts by slowing their decision making processes. Ariai and Demartini (2024) notes that legal practitioners have depended on traditional techniques methods like keyword extraction and sentence reduction when summarizing legal documents. While these techniques have been helpful to an extent, they have failed in capturing the complex semantics and context-rich nature of legal materials. More worrying, the conventional techniques are time-consuming and costly which makes them prone to inefficiencies and human error.

With the development of natural language processing (NLP) technologies, the challenges brought by conventional techniques in legal document analysis can be addressed. By definition NLP, is a branch of artificial intelligence that focusses on how humans interact with computers using natural language thus can automate and streamline the process of legal document analysis (de Oliveira & Nascimento, 2025). The most common use for NLP in law is document review and management where AI algorithms can analyze and highlight relevant information from contracts (Medium, 2022). Interestingly, NLP systems can be trained to understand, interpret and generate legal language thus profoundly impacting on legal research, decision-making and document management. Against this background, this research paper critically explores the role of advanced NLP techniques in legal document analysis. Specifically, the paper discusses the potential NLP advantages, evaluates NLP technological models, presents current applications, evaluates the impact of these technologies, examines future trends as well as challenges faced with NLP models in legal document analysis.

## **II. PROBLEM STATEMENT**

In the legal practice and research, Jagirdar et al. (2024) observed that huge volumes of complex and large legal documents present significant barrier to legal professionals who are looking to extract and comprehend data effectively. Kusabi et al. (2024) adds that the labour intensive and time consuming nature of the current manual document analysis significantly affects efficient and prompt extraction of key legal insights. Moreover, the legal documents are unique in their structure and terminology as they include intricate references, clauses, citations and conditions that call for specialised understanding. Katz et al. (2023) continues to argue that legal documents are rarely uniform and the vast volumes of data increases the challenge for professionals seeking to extract relevant information. In light of these, there is need for robust and intelligent legal document summarizer that not only handles data efficiently but also extracts the subtleties and contexts that are essential for comprehensively understanding the legal documents. Addressing these challenges calls for developing of an advanced NLP technology that can navigate through abundant legal documents, determine key elements and present concise yet comprehensive summaries and predictions.

### III. PROPOSED SOLUTION: ADVANCED NLP TECHNIQUES FOR LEGAL DOCUMENT ANALYSIS

#### 3.1 What is natural language processing and potential advantages

Agarwal (2024) defines NLP as AI component that focuses on how humans and computers interact using natural language to understand, interpret and generate human-like text in a contextually relevant and meaningful way. Katz et al. (2023) asserts that NLP breaks complex statements into smaller components, understands the context in line with rules embedded on its training data, and offers the most human-friendly output possible. The introduction of AI and NLP in the legal sector has been a game-changer for legal experts as it presents a number of potential benefits. Zhong et al. (2020) noted that NLP techniques allows lawyers and legal experts to significantly streamline their work and reduce time needed for information processing and legal document preparation. In legal analysis, Singh (2024) opines that NLP techniques facilitates the processing of large volumes of legal texts to identify and extract key information. This helps in reducing the time spent when manually reviewing documents and uncovers important patterns and relationships that may be relevant to legal processes. Ariai and Demartini (2024) adds that NLP techniques enable machines to generate texts, answer legal questions, draft regulations and simulate reasoning which significantly revolutionise legal practice and research. Figure 1 below shows the flow-chart for NLP-based legal retrieval system that can exploited by legal professionals to obtain relevant legal answers (Ning, 2022)

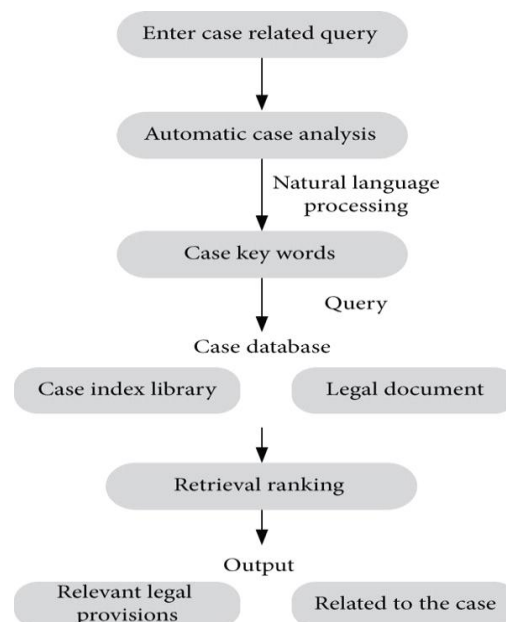


Figure 1: flow-chart for NLP-based legal retrieval system (Ning, 2022)

Recent researchers have explored the potential advantages of NLP in terms of symbol based methods which apply interpretable hand-crafted symbols to legal tasks (Surden, 2018; Zadgaonkar & Agrawal, 2021). Precisely, symbol based methods focus on utilising interpretable

legal knowledge to reason between symbols in legal documents like relationships and events. Other researchers have looked at embedding based methods that design efficient neural models to achieve better results in legal document analysis (Chalkidis & Kampas, 2019; Mukim, 2024). Precisely, embedding based methods concentrate on learning latent features for predicting legal cases from large-scale data sets. Figure 2 below shows various tasks that NLP systems can perform in legal document analysis (Zhong et al., 2020).

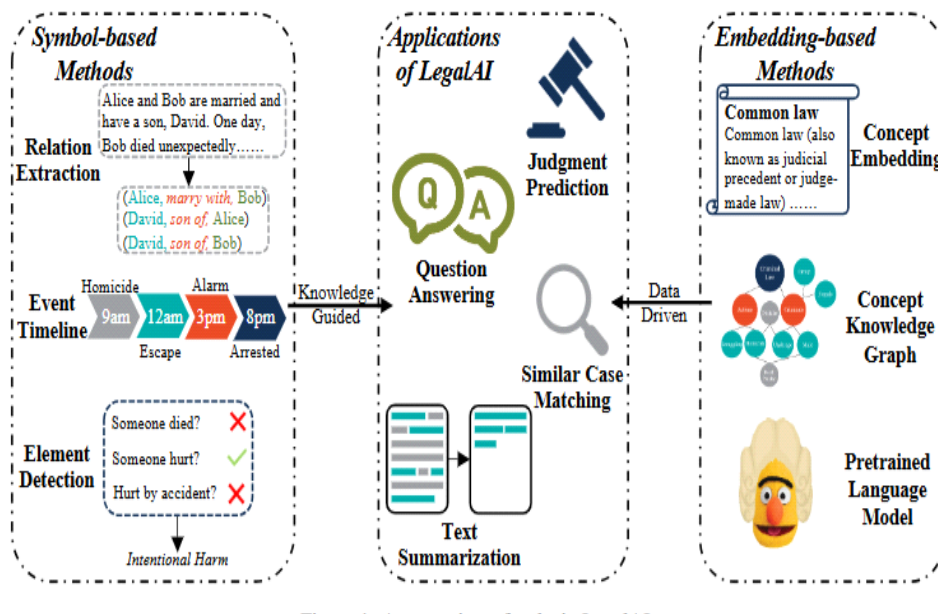


Figure 2: overview of tasks performed by NLP techniques in legal document analysis (Zhong et al., 2020)

### 3.2 Technological models and techniques in NLP for legal document analysis

Previous researchers have examined how various technological models of NLP are applied in legal document analysis. A research by Naik et al. (2023) examined named entity recognition (NER) that identifies and classifies key entities like persons, locations, laws, dates, organisations and legal roles. In legal context, NER can help in identifying parties involved in litigation, extracting referenced statutes or precedents and assisting in contract parsing to identify key obligations and parties. Ariai and Demartini (2024) highlighted that standard NER language models like Stanford NER or Spacy require legal-domain adaptation like LegalNER or BERT for NER on legal corpora to achieve acceptable accuracy in real-world legal applications. Pais et al. (2021) indicated that creation of language models like Legal-BERT and Bureau-BERTo has enhanced the performance of NER tasks by adapting to the specific legal domain. Despite advances in NER systems, limitations such as complexity of legal language and need for large annotated corpora are key challenges (Siino et al. (2025).

Scholars have also explored text classification technological models which automatically assign documents or clauses to predefined categories by filtering relevant case law based on legal issues (Bambroo& Awasthi, 2021). For instance, NLP models can classify contracts into types such as employment, lease, and service level agreements or categorise case laws into topics such as torts, criminal law, or intellectual property. Hwang et al. (2022) presented LBOX OPEN, a dataset which classifies case laws based on case name and statute prediction from the factual description of specific cases. Vatsal et al. (2023) classified legal cases by adopting BERT, Legal-BERT, RoBERTa and legalFormer to legal documents and selecting text fragments that generate credible results in line with provided metric. Tewari (2024) argues that pre-trained models like LegalPro-BERT when fine-tuned on multi-label classification tasks can significantly outperform general-purpose models like vanilla BERT for legal text classification. However, challenges remain when classifying texts as legal documents can span overlapping categories and requires annotated legal datasets for supervised learning (Siino et al., 2025).

In terms of legal document summarisation, previous researchers have explored several technological models for condensing lengthy legal documents into concise summaries to help in faster comprehension. A recent research by Licari and Comandè (2024) deployed the Italian Legal-BERT-SC model to summarise the most relevant sentences from a new dataset of ITA-CaseHold which contained more than 1,1000 pairs of judgements. They concluded that the new system outperformed other baseline models by accurately summarising the most relevant contents. Hwang et al. (2022) presented LBOX OPEN- a legal corpus having Korean precedents, as well as summarisation task comprising of supreme court precedents and the corresponding summaries. Differently, Shen et al. (2022) introduced Multi-LexSum which is a collection of 9,280 summaries generated from civil rights litigation clearinghouse and found that current NLP models such as BART, PRIMERA, LED and PEGASUS resulted in inadequate summaries as compared to expert-generated summaries. While these studies highlight the potential of NLP technological models for summarising legal documents, the findings also show that there is need to improve the summary accuracy and quality especially for complex legal documents. Singh (2024) presented techniques for automatic summarisation of legal documents using NLP models as shown in figure 3 below;

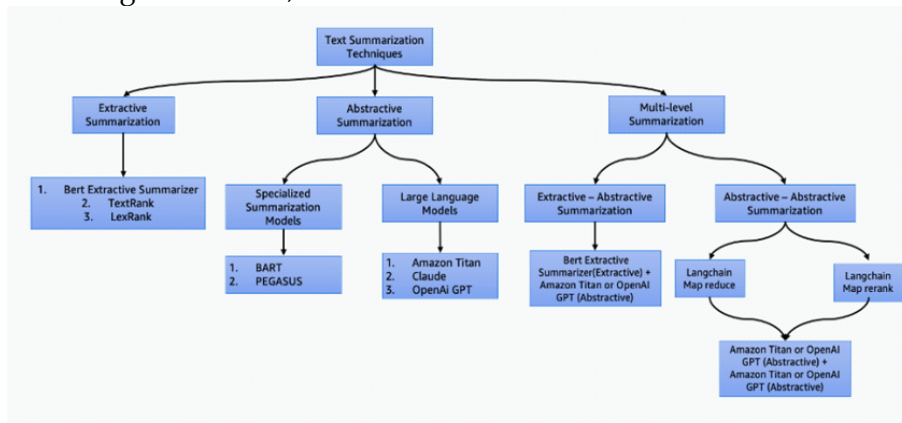


Figure 3: legal text summarisation techniques using NLP models (Singh, 2024)

With regard to judgement prediction, legal professionals have utilised models to forecast the outcome of legal cases that involves analysis of large amounts of legal data, case facts, past court decisions and applicable laws. Niklaus et al. (2023) presented multilingual dataset centred around 85,000 cases from the Swedish Supreme court. The authors deployed state of the art BERT based models specifically LongBERT, Standard BERT and Hierarchical BERT on the original case judgements and found that majority system tends to generate favourable results with regard to micro-F1 while Hierarchical BERT was a superior choice for Macro-F1 in judgement prediction. Masala et al. (2021) introduced the Romanian BERT Model that is pre-trained on large specialised corpus and claimed that the model outperformed several strong baselines for predicting legal judgements on two different corpora. Moreover, Valvoda et al. (2023) focussed on predicting cases with negative outcomes as opposed to those with positive outcomes by using two probabilistic models available on GitHub to address this challenge. The findings revealed that while basic BERT-based classification model can predict positive outcomes with a score of 75.06, it only achieved negative outcome score of 10.09, performing below random baseline that reaches a score of 11.12. This finding highlights the challenges in predicting negative outcomes in judgement prediction. Siino et al. (2025) summarised key technological models, trends and approaches in legal NLP tasks as shown in table 1 below.

**TABLE 5. Technologies and trends in legal NLP tasks**

Legal Task	Key Technologies	Trends and Approaches
<i>Document Retrieval</i>	BERT, Sentence-BERT, LEGAL-BERT, BM25, GloVe, Doc2Vec, CatBoost, TF-IDF, Longformer	Combines traditional retrieval methods with Transformers; uses similarity detection techniques.
<i>Question Answering</i>	BM25, Legal GloVe, Legal Siamese BERT	Tailored for legal queries; integrates embeddings and traditional retrieval methods to enhance performance.
<i>Named Entity Recognition (NER)</i>	BERT, LEGAL-BERT, RoBERTa, DeBERTa, Bi-LSTM, CRF, BureauBERTo, UmBERTo	Combines sequence models with Transformers; adapts models for specific legal domains.
<i>Similarity Detection</i>	BERT, LEGAL-BERT, RoBERTa, Sentence-BERT, GloVe, Doc2Vec, TF-IDF	Uses enhanced embeddings and integrates text and network-based similarity measures.
<i>Classification</i>	BERT, LEGAL-BERT, RoBERTa, DeBERTa, Longformer, SVM, CNN, LSTM, Bi-LSTM	Extensive use of Transformers; traditional classifiers for baseline comparisons.
<i>Document Summarization</i>	BERT, PEGASUS, BART, LED, PRIMERA	Applies both extractive and abstractive summarization; multitask learning approaches.
<i>Datasets and Benchmarking</i>	BERT, RoBERTa, DeBERTa, Longformer, MiniLM, DistilBERT, mDeBERTa-V3, XML-R	Focus on multilingual and multi-label datasets; evaluates pre-trained models on legal data.
<i>Document Review and Automation</i>	Vicuna, ChatGPT, Span NLI BERT, ALeaseBERT	LLMs used for contract drafting, review, and automation.
<i>Judgement Prediction</i>	BERT, LEGAL-BERT, BiGRU, RoBERTa, Logistic Regression, SVM, Random Forest, XL-Net, Longformer	Hierarchical models, domain-specific BERT models, and multilingual capabilities for predicting legal judgments.

Table 1. technological models and trends in legal NLP tasks (Siino et al., 2025)

**IV. APPLICATIONS OF NLP MODELS IN LEGAL PRACTICE**

NLP models have wide applications in the legal field. Hassan et al. (2021) pointed that NLP models play a crucial role in contract review by automatically identifying and analysing key clauses in contracts such as force majeure clauses, liability terms and confidentiality agreements. Trained on extensive collection of contract data and collective knowledge of leading legal experts, SpotDraft’s VerifyAI can process and understand intricacies of contract language thus facilitating efficient and accurate review process (Agarwal, 2024). The tool can flag potential risks, offer suggestions, provide detailed answers and help ensure compliance thus reducing the time spent on manual contract review. Apart from contract review, NLP technological models are useful in legal research where they facilitate extraction of relevant information from extensive databases, case laws and legal documents (Siino et al., 2025). Spark NLP is crucial can quickly analyse and summarize complex texts thus substantially reducing the effort and time traditionally required for exhaustive legal research (John Snow Labs, 2025). As a consequence, legal practitioners are empowered to make informed decisions on the legal landscape.

**Which NLP Libraries does your organization use?**

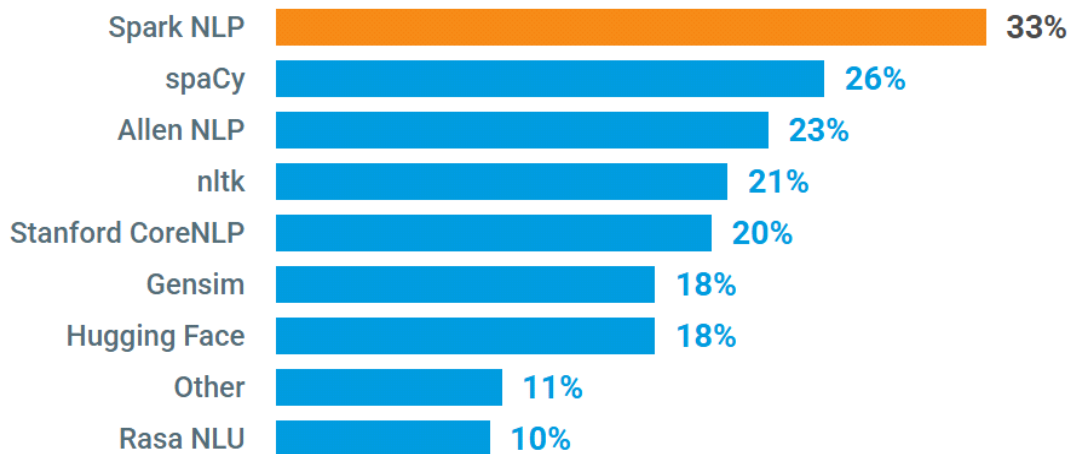


Figure 4 below summarises the NLP libraries based on organisational usage preference (John Snow Labs, 2025)

Additionally, NLP models plays a critical role in predictive legal analytics where they analyse historical legal data to identify patterns, trends and potential outcomes (Ahmed, 2023). CaseCrunch developed an AI model that predicts the outcomes of legal cases by using machine learning algorithms to analyse case data and provide predictions thus assisting lawyers to assess the case strength and develop strategies (Siino et al., 2025). Further, Imogen et al. (2024) noted that the integration of NLP-powered chatbots within legal domains has transformed the

way clients seek guidance and information. AI-powered chatbots such as Law ChatGPT, LegalMind, RightsBot and JusticeBot can democratize access to legal advice thus providing affordable options for individuals who cannot afford traditional legal services (LinkedIn, 2024). NLP-based systems can assist in interpreting legal documents and answering basic legal questions for those without legal expertise.

## **V. IMPACT ANALYSIS**

- Efficiency gains in legal practice: NLP models significantly improve the efficiency of legal research, document review and case management (Sangeetha et al., 2024). Automated tools reduce the need for time-consuming manual review and allow legal professionals to focus on higher-value tasks.
- Cost reduction and accessibility: Song et al. (2022) argues that implementation of NLP tools leads to significant cost savings especially in large-scale litigation and document-heavy processes. Through automation, legal entities can allocate resources more efficiently and offer affordable services.
- Accuracy and bias: NLP tools can enhance accuracy by identifying key terms and concepts with higher precision (Song et al., 2022). However, there are concerns about algorithmic bias especially in models trained on historical data that may reflect societal biases.
- Compliance and ethical considerations: Tekade et al. (2024) assert that legal AI systems can adhere to strict privacy laws such as the European GDPR and ensure confidentiality in client data. Moreover, ethical concerns regarding algorithmic transparency and bias should be addressed to maintain the integrity of the legal profession.

## **VI. FUTURE TRENDS AND CHALLENGES ASSOCIATED WITH NLP IN LEGAL DOCUMENT ANALYSIS**

Looking at the future trends, the development of self-learning legal NLP systems is a major future trend as it will be capable of continual learning, automatically updating their understanding of legal language, case precedents and evolving statutes without manual retraining (Monette, 2025). Although promising, these systems may evolve in unintended ways thus affecting legal reliability. Another key future trend concerns the need for legal NLP systems that are capable of operating across multiple languages and legal systems (Radhika et al., 2024). Future models should support multilingual legal corpora and enable analysis of international treaties and transnational contracts. Further, future NLP models will provide real-time legal decision support for lawyers, judges and regulators (Monette, 2025). Such systems will deliver live clause recommendations, risk scores and regulatory alerts as users draft or review documents.

Looking at the challenges, Ariai and Demartini (2025) opines that legal documents often contain sensitive client information and training NLP models on these data raises privacy and

compliance concerns under laws like GDPR, HIPAA or data sovereignty frameworks. Nagy et al. (2023) argues that undertaking federated learning and privacy-preserving NLP could be promising solution in addressing these concerns. Additionally, Siino et al. (2025) argues that the paucity of high-quality open legal datasets for benchmarking NLP systems presents challenges in addition to evaluation metrics used in general NLP that may not capture the legal significance or correctness of outputs. More worrying, deploying NLP technological models in legal field may blur lines of liability especially if an AI-powered tool gives incorrect advice (Hodge, 2023). There are also concerns about unauthorized practice of law if NLP systems provide client-facing legal analysis. This calls for NLP regulations and certification protocols to ensure accuracy and compliance with legal systems.

## VII. CONCLUSION

This paper finds that the integration of advanced NLP techniques in legal document analysis has opened transformative possibilities for modern legal practice and research. As legal professionals' grapples with the exponential growth of complex and voluminous data, NLP provides a scalable, efficient, and intelligent solution to automate tasks such as contract review, case summarization, legal research and outcome prediction. Through models like Legal-BERT, RoBERTa and domain-specific summarization systems, NLP models have demonstrated remarkable potential in reducing manual workloads, improving document comprehension, and enhancing decision-making accuracy. However, the deployment of these technologies faces challenges such as data privacy, algorithmic bias, lack of explainability, and inadequate multilingual corpora. Moreover, concerns around ethical use, unauthorized practice of law and accountability require urgent regulatory oversight. Looking forward, the future of NLP in law lies in the development of self-learning, cross-jurisdictional, and real-time intelligent systems that uphold transparency, fairness, and compliance. As such, it calls for interdisciplinary collaboration between legal experts, technologists, and policymakers to ensure the responsible and ethical integration of NLP in legal workflows.

## REFERENCES

1. Radhika, N. K. Bhasin, S. R, Y. R. Raju, K. N. V. Satyanarayana and I. I. Raj, "Optimization of Natural Language Processing Models for Multilingual Legal Document Analysis," 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), pp. 1-6, 2024.
2. A. Tewari, "LegalPro-BERT: Classification of Legal Provisions by fine-tuning BERT large language model," arXiv preprint arXiv:2404.10097, 2024.
3. A.V. Zadgaonkar and A.J. Agrawal, "An overview of information extraction techniques for legal document analysis and processing," International Journal of Electrical & Computer Engineering (2088-8708), vol. 11, no. 6, pp. 43-59, 2021.

4. B. Nagy, I. Hegedűs, N. Sándor, B. Egedi, H. Mehmood, K. Saravanan,... and A. Kiss, "Privacy-preserving Federated Learning and its application to natural language processing," *Knowledge-Based Systems*, p. 268, 2023.
5. C. Monette, "The Future of Legal Research: AI-Powered Tools and Techniques," <https://getciville.com/the-future-of-legal-research-ai-powered-tools-and-techniques/> (Accessed: 11th April 2025).
6. D. Licari and G. Comandè, "ITALIAN-LEGAL-BERT models for improving natural language processing tasks in the Italian legal domain," *Comput. Law Secur. Rev.*, vol. 52, 2024.
7. D.M., Katz, D. Hartung, L. Gerlach, A. Jana and M.J. Bommarito, "Natural language processing in the legal domain," arXiv preprint arXiv:2302.12039., 2023.
8. D. Singh, "Legal documents Text Analysis using Natural Language Processing (NLP)," 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), pp. 1302-1307, 2024.
9. D. Song, S. Gao, B. He and F. Schilder, "On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study," in *IEEE Access*, vol. 10, pp. 75835-75858, 2022.
10. F.U. Hassan, T. Le and X. Lv, "Addressing legal and contractual matters in construction using natural language processing: A critical review," *Journal of Construction Engineering and Management*, vol. 147, no. 9, 2021.
11. H. Surden, "Artificial intelligence and law: An overview," *Ga. St. UL Rev*, vol. 35, no. 1305, 2018.
12. H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu and M. Sun, "How does NLP benefit legal system: A summary of legal artificial intelligence'," <https://arxiv.org/pdf/2004.12158> (accessed: 10th April 2025)
13. I. Chalkidis, and D. Kampas, "Deep learning in law: early adaptation and legal word embeddings trained on large corpora," *Artificial Intelligence and Law*, vol. 27, no. 2, pp. 171-198, 2019.
14. Jagirdar, S. Gandage, B. Waghmare and I. Kazi, "Enhancing Legal Document Summarization Through NLP Models: A Comparative Analysis of T5, Pegasus, And BART Approaches," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 12, no.3, pp. 805-817, 2024.
15. John Snow Labs. "Natural Language Processing for Legal Documents State of the Art NLP for Legal Documents." <https://www.johnsnowlabs.com/legal-nlp/#:~:text=NLP%20for%20legal%20documents%20facilitates,and%20efficiency%20to%20the%20industry.> (Accessed: 11th April 2025)
16. J. Ning, "Natural Language Processing Technology Used in Artificial Intelligence Scene of Law for Human Behavior," *Wireless Communications and Mobile Computing*, vol. 1, no. 3, pp. 1-8, 2022.

17. J. Niklaus, V. Matoshi, P. Rani, A. Galassi, M. Stürmer and I. Chalkidis, "LEXTREME: A multi-lingual and multi-task benchmark for the legal domain", Proc. Findings Assoc. Comput. Linguistics EMNLP, pp. 3016-3054, 2023.
18. J. Valvoda, R. Cotterell and S. Teufel, "On the role of negative precedent in legal outcome prediction", Proc. Trans. Assoc. Comput. Linguistics, vol. 11, pp. 34-48, 2023.
19. K. Agarwal, "Exploring NLP in Legal Practice: Use Cases, Pros, Cons & More." <https://www.spotdraft.com/blog/exploring-nlp-in-legal-practice#:~:text=By%20automating%20the%20labor%2Dintensive,and%20contextually%20relevant%20legal%20language.> (Accessed: 10th April 2025)
20. LinkedIn. "The List of Free Legal AI Chatbots in the Modern Legal Practice." <https://www.linkedin.com/pulse/list-free-legal-ai-chatbots-modern-practice-crmone-global-8rzif/> (Accessed: 11th April 2025)
21. M. Masala, R. C. A. Iacob, A. S. Uban, M. Cidota, H. Velicu, T. Rebedea, et al., "JurBERT: A Romanian BERT model for legal judgement prediction", Proc. Natural Legal Lang. Process. Workshop, pp. 86-94, 2021.
22. Medium. "Is Natural Language Processing Advanced Enough to Tackle Legal Documentation?" <https://odsc.medium.com/is-natural-language-processing-advanced-enough-to-tackle-legal-documentation-f218f9772e3d> (Accessed: 10th April 2025)
23. M. Siino, M. Falco, D. Croce and P. Rosso, "Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches," IEEE Access, vol. 13, pp. 18253-18276, 2025.
24. P. Bamroo and A. Awasthi, "LegalDB: Long DistilBERT for legal document classification", Proc. Int. Conf. Adv. Electr. Comput. Commun. Sustain. Technol. (ICAECT), pp. 1-4, 2021.
25. P. Tekade, N. Jadhav, S. Chawda, V. Bhosale and Y. Akashe, "Legal lens: Exploring NLP for Document Analysis in Law," 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Pune, India, pp. 1-5, 2024.
26. P.V. Imogen, J. Sreenidhi and V. Nivedha, "AI-Powered Legal Documentation Assistant", Journal of Artificial Intelligence, vol. 6, no. 2, pp. 210-226, 2024.
27. R. Singh, "Large Language Models and Text Summarization: A Powerful Combination". [https://medium.com/@singhrajni2210/large-language-models-and-text-summarization-a-powerful-combination-6400e7643b70.](https://medium.com/@singhrajni2210/large-language-models-and-text-summarization-a-powerful-combination-6400e7643b70) (Accessed: 11th April 2025)
28. R.S. de Oliveira and E.G.S. Nascimento, "Analysing similarities between legal court documents using natural language processing approaches based on Transformers," PLoS ONE, vol. 20, no. 4, 2025.
29. S.D. Hodge, "Revolutionizing Justice: Unleashing the Power of Artificial Intelligence," SMU Sci. & Tech. L. Rev., vol. 26, no. 217, 2023.
30. S.J. Ahmed, "From Text to Verdict: Unleashing NLP's Potential in Legal Case Prediction," <https://medium.com/dataduniya/from-text-to-verdict-unleashing-nlps-potential-in-legal-case-prediction-64c2c5f20fe4> (Accessed: 11th April 2025).

31. S. Vatsal, A. Meyers and J. E. Ortega, "Classification of U.S. supreme court cases using BERT-based techniques", Proc. Conf. Recent Advances in Natural Lang. Processing- Large Language Models, pp. 1207-1215, 2023.
32. T.L. Mukim, "Enhancing Legal Guidance by Utilizing Natural Language Processing-Based Document Embedding," Doctoral dissertation, Dublin, National College of Ireland, 2024.
33. V. Kusabi, P. Jain, R. Palesha, K. Khandelwal, G.P. Potdar & P.T. Kohok, "Advancements in Legal Document Processing and Summarization," Computer Research and Development, vol. 24, no. 6, pp. 1-10, 2024.
34. V. Naik, P. Patel and R. Kannan, "Legal entity extraction: An experimental study of ner approach for legal documents," International Journal of Advanced Computer Science and Applications, vol. 14, no. 3, pp. 775-783, 2023.
35. V. Pais, M. Mitrofan, C. L. Gasan, V. Coneschi and A. Ianov, "Named entity recognition in the Romanian legal domain", Proc. Natural Legal Lang. Process. Workshop, pp. 9-18, 2021.
36. V. Sangeetha, I. Pani, J. Jariwala, A. Dargar, R. H. Rai and S. Agarwala, "Optimizing Judicial Efficiency- A Software Solution for Case Listing with Customized Case Flow Management," 2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS), pp. 1-6, 2024.
37. W. Hwang, D. Lee, K. Cho, H. Lee and M. Seo, "A multi-task benchmark for Korean legal language understanding and judgement prediction", Proc. Adv. Neural Inf. Process. Syst., pp. 1-15, 2022.
38. Z. Shen, K. Lo, L. Yu, N. Dahlberg, M. Schlanger and D. Downey, "Multi-LexSum: Real-world summaries of civil rights lawsuits at multiple granularities", Proc. Adv. Neural Inf. Process. Syst., pp. 13158-13173, 2022.