Journal of Trends and Challenges in Artificial Intelligence

Vol. 03, No. 01 (2026) 59-66, DOI: 10.61552/JAI.2026.01.006 - http://jai.aspur.rs

LEGALEASE : AI-POWERED SUMMARISATION AND VISUALIZATION OF LEGAL DOCUMENTS

Siddhi Bhogale [1]
Nidhi Nayak
Shivani Nikam
Pranathi Narsupalli
Vidya Pujari

Received 21.02.2025.
Revised 03.05.2025.
Accepted 18.06.2025.

Keywords:

Legal document summarization, T5-small, Natural Language Processing, Abstractive summarization, Large Language Model.

Original research

A B S T R A C T

Legal documents are often complex and filled with specialized jargon, making them difficult to comprehend. LegalEase is an AI-driven tool designed for effective summarization and visualization of legal texts using mind maps. It fine-tunes the T5-small model on datasets like IN-ABS and MILDSum to generate concise, contextually aware summaries. Additionally, it utilizes the LlaMA model for structured formatting, D3.js for flowchart-based visualizations, and Google Cloud API for multilingual translation, including Hindi. The system also features jargon simplification using a JSON-based dictionary to enhance readability. Experimental results show significant improvements in summarization quality, with enhanced ROUGE, Meteor, and BERTScore metrics compared to pre-trained models. By integrating visual and multilingual tools, LegalEase improves accessibility and usability across various legal domains and audiences.

1. INTRODUCTION

Legal documents play a crucial role in governance, business, and personal affairs, yet they are often difficult to comprehend due to their length, complexity, and technical jargon. A study by the World Justice Project (2021) found that nearly 57% of individuals struggle to understand legal language, limiting their ability to make informed decisions (Wszalek, 2017). Moreover, lawyers and legal professionals spend up to 30% of their time reviewing lengthy case files, contracts, and regulatory documents, as reported by Thomson Reuters' 2022 Legal Trends Report (Widuhung & Machmud, 2024). The increasing volume of legal documents combined with the time constraints faced by professionals, highlights the need for efficient legal summarization tools.

Traditional methods of summarizing legal texts are time-intensive and prone to errors. While AI-powered solutions exist, many fail to preserve the original meaning or structure of legal arguments. Generic AI models often struggle with legal-specific terminology, leading to misinterpretations that could have serious consequences in legal proceedings. A clear example of this is seen in regulatory compliance, where misinterpreted clauses can result in financial or legal penalties for businesses. The lack of accessible legal information further widens the gap between legal professionals and individuals who need legal assistance but cannot afford expert consultations.

[1] Corresponding author: Siddhi Bhogale
Email: 2021.siddhi.bhogale@ves.ac.in

To address these challenges, this paper introduces LegalEase, an AI-powered system designed to generate structured, accurate, and accessible summaries of legal documents. The system fine-tunes the T5-small model on legal datasets such as IN-ABS and MILDSum to produce more contextually accurate summaries. Additionally, it simplifies complex legal jargon by identifying and providing simplified explanations for key terms, improving readability for non-experts. The system also improves comprehension through visual flow charts generated using D3.js, helping users grasp key legal points in a structured manner. Furthermore, it incorporates multilingual translation using Google Cloud API, making legal summaries accessible to a wider audience, including non- English speakers.

With these features, LegalEase aims to improve the accessibility of legal documents for professionals, researchers, and individuals who need quick and accurate legal insight. This paper explores the development, implementation and evaluation of the system, demonstrating how AI can bridge the gap between complex legal texts and user-friendly summaries.

2. LITERATURE REVIEW

Legal document summarization has been an active area of research, with various methodologies that address challenges such as domain-specific text processing, semantic understanding, and extractive versus abstractive summarization. Previous works have primarily focused on sentence scoring, chunk- based summarization, word embeddings, and transformer- based models for legal texts.

Polsley et al. (2016) proposed a sentence scoring approach that ranks sentences based on their relevance in legal documents. Their method leverages TF-IDF weighting and POS tagging to identify abbreviations and key terms specific to the legal domain. Although effective, this method misses important sentences and struggles with inefficient section representation. A domain-specific scoring mechanism can enhance the understanding of legal documents based on flowcharts.

Yash et al. (2018) explored word comparison in the Indian judiciary using Singular Value Decomposition (SVD) and cosine similarity within a Latent Semantic Analysis (LSA) framework. Their approach effectively identifies word-level semantic similarities, though it is limited to word frequencies rather than capturing conceptual relationships. This highlights the need for semantic analysis techniques that improve translation accuracy and deeper understanding of legal terms.

Paheli et al. (2021) introduced DELSumm, an extractive summarization method focusing on content words while sum- marizing Supreme Court judgments. Their approach efficiently extracts key sentences, but struggles with certain rhetorical roles, leading to information gaps. A combination of extractive techniques with improved relevance detection can improve legal text summarization.

Feijo et al. (2023) LegalSumm proposed chunk-based summariza- tion, dividing Portuguese legal documents into independent sections, summarizing each chunk separately, and combining them into a final summary. Although effective for structured legal texts, context loss between chunks remains a challenge. This approach suggests chunk-based summarization as a potential technique for summarizing large legal documents in multiple languages.

Recent advancements in abstractive summarization have leveraged transfer learning to handle complex legal texts. Zhang and Alvarado-Barahona (2023) applied Big Bird, a transformer model capable of processing long legal documents (up to 4096 tokens). However, high computational costs remain a concern, necessitating a balance between efficiency and accuracy.

Chalkidis et al. (2024) introduced LexGLUE, a bench- mark dataset for understanding legal language, which evalu- ates transformers such as BERT and RoBERTa in classification and prediction of case outcomes. However, LexGLUE does not directly address summarization, indicating a potential research direction in which legal summarization models could incorporate additional LexGLUE datasets.

Although prior research has contributed to extractive (Shukla et al., 2022), abstractive, and chunk-based summarization, there remains a gap in efficiently condensing legal texts while maintaining factual consistency. Existing methods either suffer from information loss (extractive models) or high computational costs (transformer-based models). Our approach aims to leverage T5-small, fine-tuned on IN-ABS (Bhattacharya et al., 2019) and MILDSum (Datta et al., 2023) datasets, to improve legal text condensation while balancing efficiency and accuracy.

Sasidharan and Rahulnath (2023) proposed a structured approach for relation extraction in legal documents, utilizing knowledge graphs to represent entities and their relationships. Their method integrates graph reasoning and clustering techniques to enhance legal concept organization and retrieval. By structuring legal knowledge as a graph, the study facilitates better legal research and decision-making. However, the paper does not focus on summarization, suggesting a potential research direction where mind maps or graph-based summarization techniques could be incorporated to improve legal text condensation and visualization.

Incorporating multilingual support (Gupta & Mehta, 2023) into legal document summarization systems is essential for enhancing accessibility and ensuring equitable access to justice in linguistically diverse regions. For instance, in India, where legal judgments are pre- dominantly authored in complex English, a significant portion of the population may find it challenging to comprehend these documents. Translating summaries into regional languages, such as Hindi, can bridge this gap, enabling a broader audience to understand legal outcomes and their implications.

3. METHODOLOGY

3.1 Legal Document Summarization

To produce accurate and coherent legal summaries, the T5- small model was fine-tuned with IN-ABS and MILDSum, two datasets built exclusively for legal text summarization. Given the hierarchical and intricate nature of legal papers, preserving contextual flow across parts was a big problem.

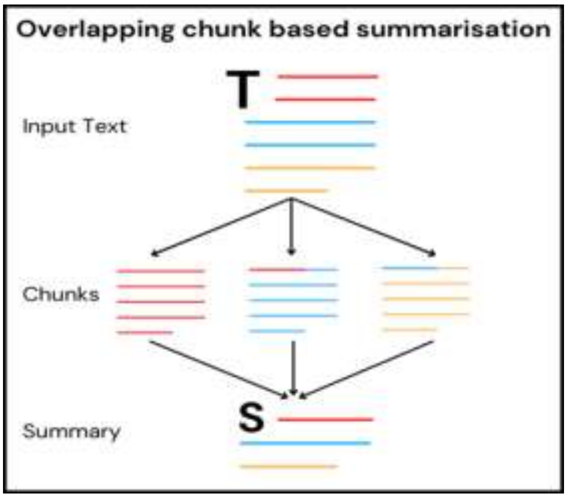

Figure 1. Overlapping chunk based summarization

To overcome this, an overlapping chunk-based abstractive summarization technique was used, with each input segment or chunk overlapping by 150 tokens (Figure 1). This method helps to maintain context across chunks within the text, avoiding abrupt transitions and missing essential references between portions. By utilizing this strategy, the model increases the continuity and coherence of generated summaries, making them more trustworthy for legal professionals and researchers.

3.2 BNS Legal Search

A Flask-based backend is used to implement the BNS Legal Search Feature, providing the framework for managing user requests and query processing. Sentence Transformers (paraphrase-MiniLM-L6-v2) are used in natural language pro- cessing to provide vector embeddings for user queries and BNS descriptions, guaranteeing a semantic understanding of legal materials (Figure 2).

Text-based similarity matching is implemented by TF-IDF and Cosine Similarity, which enables the system to rank and extract the most relevant legal portion. The legal data, which includes BNS sections, offenses, descriptions, and punishments, is kept in an organized dataset based on CSV. There are three ways for users to look for relevant legal sections:

- Offense-based search;
- Section-based search;
- Query-based search.

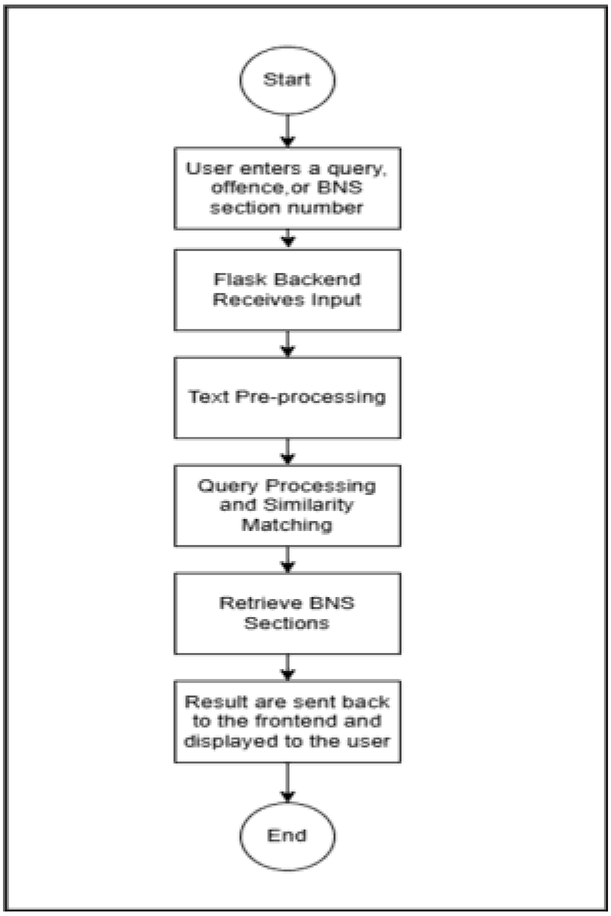

Figure 2. BNS Legal Search System Workflow

3.3 Jargon Simplification

The Legal Jargon Identification function in LegalEase helps to clarify legal summaries by identifying and defining complex legal terms (Garimella et al., 2022). In legal documents, technical jargon that may prove difficult for non-experts is commonplace (Gallegos & George, 2022). To address this, LegalEase examines the summary generated using Natural Language Processing (NLP) techniques, identifying areas where difficult legal terms occur.

These terms are then cross-referenced with a pre-built JSON dictionary, which contains definitions of complex legal words. Such terms are highlighted in turn with users presented on mouse hovers with short explanatory texts. This feature thus alleviates the burden of consulting external references for quick understanding of legal jargon by the user (Cemri et al., 2022).

This feature is scalable and adaptable. Updating a dictio- nary and even providing multilingual support are a few of the options handy. Legalease enhances accessibility through automated term detection against a structured legal dictionary and an interactive hover-based explanation feature, hence promoting transparency and user-friendliness in legal content.

3.4 Mind map based visualisation for Legal document comprehension

Legal papers can have a complicated hierarchical structure that makes it challenging for consumers to effectively understand important clauses and connections. In order to solve this, LegalEase uses D3.js to integrate a mind map- based visualization system, which makes legal documents more readable and allows for structured representation.

The summarization model first extracts key sections, provi- sions, and clauses from the legal document. These components are then structured into a hierarchical JSON format, where each node represents a legal concept, and edges define relationships between clauses. The root node represents the over- arching legal topic, with child nodes branching into relevant subsections.

The system employs D3.js to generate an interactive node- link diagram, allowing users to navigate the document visually.

4. IMPLEMENTATION

4.1 Technology Stack

The system is built using a combination of machine learning models, natural language processing techniques, and visualization tools to enhance legal document summarization and accessibility (Figure 3). The key technologies used include:

- T5-small: Employed for summarization of legal documents.
- D3.js: Used for generating interactive mind maps.
- Google Translate API: Provides multilingual support for legal summaries.
- JSON-based legal dictionary: Facilitates jargon simplification by defining complex legal terms.

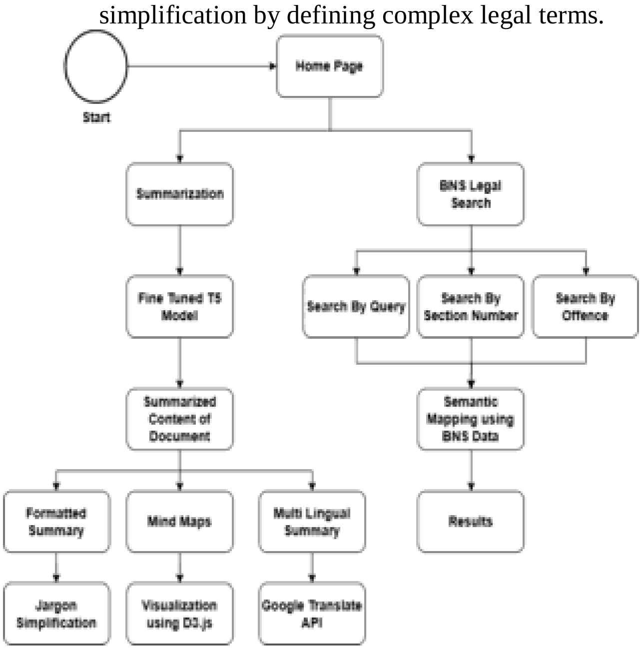

Figure 3. Activity Diagram

4.2 UI Implementation of hover-based term definitions

To enhance user experience, the system implements hover- based term definitions, allowing interactive and seamless access to legal explanations. When a summary is generated, com- plex legal terms are automatically highlighted and linked to a prebuilt JSON dictionary containing their definitions. The text is processed to wrap the legal terms with appropriate HTML elements, such as <span> tags with a data-tooltip attribute. This processed HTML is then rendered with CSS handling tooltip behavior.

Using CSS hover effects, tooltips dynamically appear when users hover over a term, displaying its definition in a clear and readable format. This interactive design ensures that users can access legal term definitions instantly without external searches. Additionally, the UI remains fully responsive, maintaining compatibility across various devices and screen sizes.

4.3 Fine-tuning of T5-Small and dataset preprocessing

To generate concise and context-aware summaries, we fine- tuned the T5-small model on two legal summarization datasets: IN-ABS (Bhattacharya et al. 2019) and MILDSum (Datta et al., 2023). These datasets were combined, resulting in a total of approximately 8,200 annotated legal examples (Legal documents and their respective summaries) for training and evaluation.

The dataset preprocessing involved removing redundancies, standardizing legal terminology, and handling special characters to improve model efficiency. To further enhance generalization, data augmentation techniques such as paraphrasing and sentence shuffling were applied. These enhancements allowed the model to generate more reliable and structured summaries while preserving key legal insights.

The training and evaluation datasets were tokenized using the T5 tokenizer, with truncation and padding applied to both inputs and targets. The inputs were prefixed with the ”summarize:” instruction to guide the model’s behavior, as per T5’s design.

4.4 Overlapping Chunk-based summarization

Legal documents are often lengthy, which poses a challenge for transformer models due to token limitations. To overcome this, we implemented an overlapping chunk-based approach, where each chunk is 500 tokens long and each input segment overlaps by 100 tokens. This technique ensures that contextual continuity is preserved across segments, reducing information loss and enhancing the coherence of generated summaries.

4.5 BNS Legal Search

To ensure semantic hold of the text, the system generates vector embeddings for user questions and BNS descriptions using Sentence Transformers. To improve the retrieval accuracy, TF-IDF similarity is used to give additional textual similarity matching. A Flask-based API that incorporates the entire system provides an intuitive user interface for smooth communication. To ensure complete legal coverage, the data set is made up of an organized collection of BNS sections, descriptions, offenses, and repercussions.

4.6 Translation

Using the Google Translate API, the system integrates multilingual capabilities to make legal summaries available to a larger audience. Users can choose their preferred language after a summary is prepared, and the system will translate the material dynamically while maintaining its original structure. By ensuring that legal terminology is precise and appropriate for the context, the translation process reduces the possibility of misunderstandings. This function expands the reach of legal information and improves accessibility for non-native English speakers.

4.7 Mind Map

The system leverages D3.js to generate interactive mind maps that visually structure key paragraphs and concepts in legal documents, enhancing comprehension. By organizing information hierarchically, this approach enables users to quickly grasp the relationships between different legal ideas.

The summarized text is segmented into nodes and edges, where each node represents a key legal concept, and edges define the relationships between clauses and phrases. This structured visualization improves readability, reduces cognitive load, and allows both legal professionals and non-experts to efficiently analyze complex legal texts. The interactive design further enhances engagement, enabling users to explore legal information dynamically, leading to faster and more effective comprehension.

5. RESULTS AND EVALUATION

5.1 Comparative performance of T5-Small and Fine-tuned T5-small

To ensure a robust evaluation, the model was assessed using both lexical and semantic similarity metrics.

Table 1. Comparison of T5-small and Fine-tuned T5-small performance

Metric	T5-small	Fine-tuned T5-small
ROUGE-1	0.0201	0.0303
ROUGE-2	0.0000	0.0000
ROUGE-L	0.0201	0.0242
METEOR Score	0.0386	0.0555
BERT F1 Score	0.7716	0.7729
BERT Precision	0.7081	0.7161
BERT Recall	0.8475	0.8395

The ROUGE scores measured n-gram overlaps, while BERTScore assessed contextual alignment. BLEU and METEOR further validated fluency and accuracy. This multi-metric approach enabled a well-rounded assessment of summary quality across legal content.

The comparative evaluation of the original T5-small model and the fine-tuned T5-small model based on key summarization metrics is presented in Table I.

5.2 Parameters for Fine-tuning

The best-performing fine-tuning configuration was determined empirically after experimenting with several combinations of hyperparameters. The final setup was as follows:

- Learning Rate: 2e-5
- Epochs: 4
- Training Batch Size: 8
- Evaluation Batch Size: 4
- Weight Decay: 0.01
- Evaluation Strategy: per epoch

The fine-tuning process was executed using Hugging Face’s Seq2SeqTrainer API. The model demonstrated consistent improvement across epochs and was subsequently evaluated using multiple metrics including ROUGE, BERTScore, BLEU, and METEOR, providing a comprehensive view of summarization quality.

5.3 Analysis of Results of Summarization

The fine-tuned T5-small model demonstrated superior performance compared to the base T5-small model across multiple evaluation metrics. The ROUGE-1 score increased from 0.0201 to 0.0303, indicating a higher degree of overlap between generated and reference summaries. Additionally, a modest improvement in the BERT F1-score suggests better balance between precision and recall.

The METEOR score showed a significant rise from 0.0386 to 0.0555, reflecting improved synonym matching and gram- matical fluency in the generated summaries. BERT-based evaluations further indicate that the fine-tuned model achieved a precision score of 0.7161 and a recall score of 0.8395, highlighting its ability to generate semantically meaningful and contextually rich summaries.

The use of an overlapping chunk-based approach contributed to better retention of context, resulting in more coher- ent legal summaries. However, the ROUGE-2 score remained at 0.0000, likely due to the inherent complexity of legal terminology and sentence structures, indicating potential areas for further optimization. Overall, these results demonstrate that fine-tuning on legal datasets significantly enhances summarization performance, making LegalEase more effective for legal document processing.

6. DISCUSSION

6.1 Strengths of LegalEase compared to other Legal AI tools

LegalEase distinguishes itself from other legal AI tools by offering a comprehensive suite of features, including fine- tuned summarization, jargon simplification, mind map generation, and crime classification. The integration of these capabilities within a single platform provides a holistic approach to legal document analysis, making it versatile for both legal professionals and non-experts (Ladhak et al., 2023)

Additionally, the implementation of overlapping chunk-based summarization ensures that long documents are processed while maintaining context, overcoming a key limitation of traditional summarization models that struggle with lengthy legal texts. This capability enhances the accuracy and coher- ence of generated summaries, making LegalEase a robust tool for legal document analysis.

6.2 Challenges in handling legal ambiguity

One of the primary challenges for LegalEase is addressing the inherent ambiguity in legal language (Kumar & Sharma, 2023). Legal documents often contain nuanced terminology, complex sentence structures, and multi-layered interpretations that may vary by jurisdiction. While the system performs well with structured data and standard legal terminology, it may struggle with ambiguous or context-dependent language, potentially affecting the reliability of summaries and crime classifications.

6.3 Ethical Considerations

The use of AI in legal applications raises ethical concerns, particularly regarding the accuracy and reliability of automated legal insights. Although LegalEase is designed to support legal professionals, there remains a risk of providing incomplete or misleading information if the model fails to capture legal nuances or recent legislative updates.

Ensuring the ethical use of AI-driven legal tools requires:

- Continuous monitoring and refinement of model performance
- Transparency in the decision-making process of the AI system.
- Human oversight to validate generated summaries, particularly in complex legal matters that have significant implications for individuals and society.

7. FUTURE WORK

Future enhancements for LegalEase will focus on improving summarization quality by incorporating additional legal datasets (Zhong et al., 2022), allowing the model to generalize across a broader range of legal texts. Another key area for development is dynamic jargon expansion, where machine learning models generate real-time definitions for legal terms based on context, further improving accessibility for non-experts.

Additionally, the deployment of LegalEase as a browser extension or interactive web-based tool is under consideration, enabling legal professionals and the general public to access its summarization and jargon simplification features seamlessly within their workflows. These advancements will further enhance usability, making legal information more accessible and interpretable.

8. CONCLUSION

This study introduced LegalEase, an AI-powered legal assistant that enhances the comprehension of legal documents through abstractive summarization, jargon simplification, mind map generation, crime classification, and multilingual translation. The key contributions of this system include:

- Fine-tuning a T5-small model for legal text summarization.
- Implementing an overlapping chunk-based abstractive summarization technique to retain context across text segments.
- Providing real-time jargon explanations through a JSON- based dictionary.

The potential impact of LegalEase is significant for lawyers, researchers, and the general public. By streamlining legal document analysis, improving readability, and increasing accessibility to legal knowledge, the system serves as a valuable tool, particularly for non-experts.

As AI continues to evolve, its role in legal document processing is expected to expand further, automating repetitive tasks, delivering deeper insights, and optimizing legal workflows. With continued advancements in machine learning and natural language processing, systems like LegalEase have the potential to become indispensable in modern legal prac- tice, ensuring greater transparency and accessibility to legal information.

References

Bhattacharya, P., Hiware, K., Rajgaria, S., Pochhi, N., Ghosh, K., & Ghosh, S. (2019). A comparative study of summarization algorithms applied to legal case judgments. In European Conference on Information Retrieval (pp. 413–428). Springer.

Cemri, M., Çukur, T., & Koç, A. (2022). Unsupervised simplification of legal texts. arXiv preprint, *arXiv*:2209.00557.

Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D. M., & Aletras, N. (2024). LexGLUE: A benchmark dataset for legal language understanding in English. arXiv preprint, *arXiv*:2407.00648.

Datta, D., Soni, S., Mukherjee, R., & Ghosh, S. (2023). MILDSum: A Novel Benchmark Dataset for Multilingual Summarization of Indian Legal Case Judgments. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 5291–5302). Association for Computational Linguistics.

Feijo, D., Silva, T., & Pereira, L. (2023). LegalSumm: Chunk-based summarization of Portuguese judgments. *Expert Systems with Applications*, 213, 119086.

Gallegos, I., & George, K. (2022). The right to remain plain: Summarization and simplification of legal documents. Stanford CS224N Custom Project Report. Stanford University.

Garimella, A., Sancheti, A., Aggarwal, V., Ganesh, A., Chhaya, N., & Kambhatla, N. (2022). Text simplification for legal domain: Insights and challenges. Proceedings of the Natural Legal Language Processing Workshop 2022, 28. Association for Computational Linguistics.

Gupta, A., & Mehta, S. (2023). Natural language processing in legal document summarization: A review. *International Education and Research Journal*, 9(6), 45–52.

Kumar, R., & Sharma, P. (2023). AI-driven legal text summarization: Challenges and opportunities. *Engineered Science*, 10(22), 155–172.

Ladhak, F., Durmus, E., Cardie, C., & McKeown, K. (2023). Exploring legal document summarization with pretrained transformers. arXiv preprint, *arXiv*:2310.18600.

Paheli, S., Gupta, N., & Sharma, A. (2021). DELSumm: Extractive summarization of informative sentences with content word emphasis. *Journal of Data Science and Analytics*, 29(3), 203–220.

Polsley, R., O'Hare, M., & McDonald, S. (2016). Sentence scoring and domain processing in legal document summarization. *Journal of Legal Information*, 34(2), 123–145.

Sasidharan, A. K., & Rahulnath, R. (2023). Structured approach for relation extraction in legal documents. In 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India.

Shukla, A., Bhattacharya, P., Poddar, S., Mukherjee, R., Ghosh, K., Goyal, P., & Ghosh, S. (2022). Legal case document summarization: Extractive and abstractive methods and their evaluation. Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing.

Widuhung, S. D., & Machmud, A. (2024). The effect of earning per share, environment and social disclosure on company performance: case study of non-financial companies listed in Reuter and Thomson (2018-2022). *Journal of Information Systems Engineering and Management*, *10*(17s), 465-475.

Wszalek, J. (2017). Ethical and legal concerns associated with the comprehension of legal language and concepts. *AJOB neuroscience*, *8*(1), 26-36.

Yash, S., Kumar, R., & Patel, A. (2018). Word comparison using SVD and cosine similarity with LSA in the Indian judiciary domain. *International Journal of Computational Linguistics*, 45(4), 456–472.

Zhang, M., & Alvarado-Barahona, A. (2023). Abstractive summarization of legal text corpuses using transfer learning. Stanford CS224N Custom Project.

Zhong, Z., Zeng, Y., Cui, L., & Zhao, W. X. (2022). A dataset for legal case summarization. Proceedings of ACL 2022, 4293–4304.

Siddhi Bhogale
VESIT, Vivekanand Education Society's Institute of Technology
India.
2021.siddhi.bhogale@ves.ac.in
ORCID: 0009-0004-0452-8168

Nidhi Nayak
VESIT, Vivekanand Education Society's Institute of Technology
India
2021.nidhi.nayak@ves.ac.in
ORCID: 0009-0009-0151-5499

Shivani Nikam
VESIT, Vivekanand Education Society's Institute of Technology
India
2021.shivani.nikam@ves.ac.in
ORCID: 0009-0008-7177-1992

Pranathi Narsupalli
VESIT, Vivekanand Education Society's Institute of Technology
India
2021.pranathi.narsupalli@ves.ac.in
ORCID: 0009-0001-6156-1114

Vidya Pujari
VESIT, Vivekanand Education Society's Institute of Technology
India
vidya.pujari@ves.ac.in
ORCID: 0000-0002-8023-7379