**LawSage.AI**

Abhinav Srivastava Parth Shandilya Vikrant Sharma

E22CSEU0844 E22CSEU0856 E22CSEU907

**Abstract**

Legal texts are very complex texts for a common man. They are time-consuming to and require going through a vast number of documents, books, or cases to understand the jargon terms as needed. As the era of large language models is developing day by day for various fields such as text generation, code generation, or solving mathematics, we don’t see any specific dedicated LLM trained for legal research. There exist some models, but they are limited to one or a couple of countries legal data and not trained for a world-level use or specific country use. In this project, we aim to create an AI-powered web application that will have two main features: first, a legal research assistant that will provide answers and insights specifically to legal queries, and second, a document summarization and simplification feature that will extract concise summaries of lengthy legal documents and simplify legal jargon for better understanding. Our whole work is focused on the Indian legal context. Later in this report, we have shared the model details and web application structure and respected details.

**Introduction**

In the field of law, a significant amount of written content is essential such, as laws court rulings contracts and other documents Legal opinions can be quite detailed and intricate since they frequently incorporate terminology that can pose a challenge, for experts, in the field. Ordinary people alike are impacted by the system in various ways – from lawyers and judges, to legal assistants and aspiring law students. People often come across deadlines that require them to work precisely. Large volumes of data are involved in cases where concise and straightforward summaries of paperwork are needed. are incredibly beneficial, in the realm of Legal Document Summarization and Simplification (LDSS) which's a growing area of interest. LDSS tackles this issue by using cutting edge technology, in Natural Language Processing (NLP). There are two approaches, to summarization. Extractive and abstractive. Creating a summary entail condensing information whereas the abstractive approach involves formulating sentences that convey the essence of the content. Selecting sentences or phrases directly from the text is part of the extractive process. Condensing texts, into more concise forms is a common practice in the legal document summarization system (LDSS), without making any alterations to the original content. Creating coherent structures to maintain information is key even if abstracting offers more opportunities, in communication. Facing difficulties may be tough; however this method has the benefit of creating smooth summaries that mimic language.

In conversations, among people or in writing, for the audience simplification involves making intricate legal terms easier to understand by replacing terms and intricate sentence structures with simpler ones the use of expressions helps to make the content easier for everyone to understand and connect with on a scale because of the combined effect of The combined approach of summarizing and simplifying leads to a method that produces more digestible content. Simplified versions of papers, for understanding. Lately, in the field of natural language processing (NLP) there have been advancements, with the rise of transformer-based structures.

Models, like BERT have had an impact, on improving the ability to condense and simplify documents. Bidirectional Encoder Representations, from Transformers and GPT (Generative Pretrained Transformer). Text To text Transfer Transformer (referred to as T55 and Pegasus) have demonstrated achievements, in text related tasks due, to their advanced capabilities. Due, to their enhanced language processing skills these advancements have made it possible to develop specialized adjustments, like LegalBERT that's a customized version of BERT designed specifically for purposes. To documents alteration, by Legal Pegasus is the adaptation of the Pegasus model to improve summarization purposes. Extensive legal documents are their specialty these models excel at handling text and extracting information within them. This study examines how the latest transformer models are being used in the field of technology. LDSS is being developed to automatically generate summaries and simplified versions of documents. Legal documents are known for their complexity due, to the use of terminology and formal writing style. A strong LDSS should balance using citations and references, from sources effectively. Striking a balance, between upholding the authenticity of the information and presenting it in an understandable manner. The aim of this research is to evaluate the effectiveness of transformer-based models. Legal Pegasus and BERT contribute to striking a balance by generating summaries that're comprehensive and concise, easy to understand.

Moreover we evaluate the effectiveness of these models using established benchmarks. Indicators, like ROUGE and BLEu scores are commonly used to assess how content matches and the quality of translations. In turn we can measure the success of our endeavors accurately and also by explaining ideas LDSS systems can help connect people who may not have knowledge. The spread of knowledge, among the public and lawyers has led to an accessible understanding of the law. Empower individuals, by increasing their understanding of their rights and duties to enable them to make decisions. Making informed decisions without relying heavily on expert guidance is crucial, in various scenarios involving the analysis of legal documents structure and content. We aim to create a solution, for summarizing decisions and laws incorporating court judgments and statutes into our work. Streamlining documents could potentially drive progress in the field of technology and boost its development.

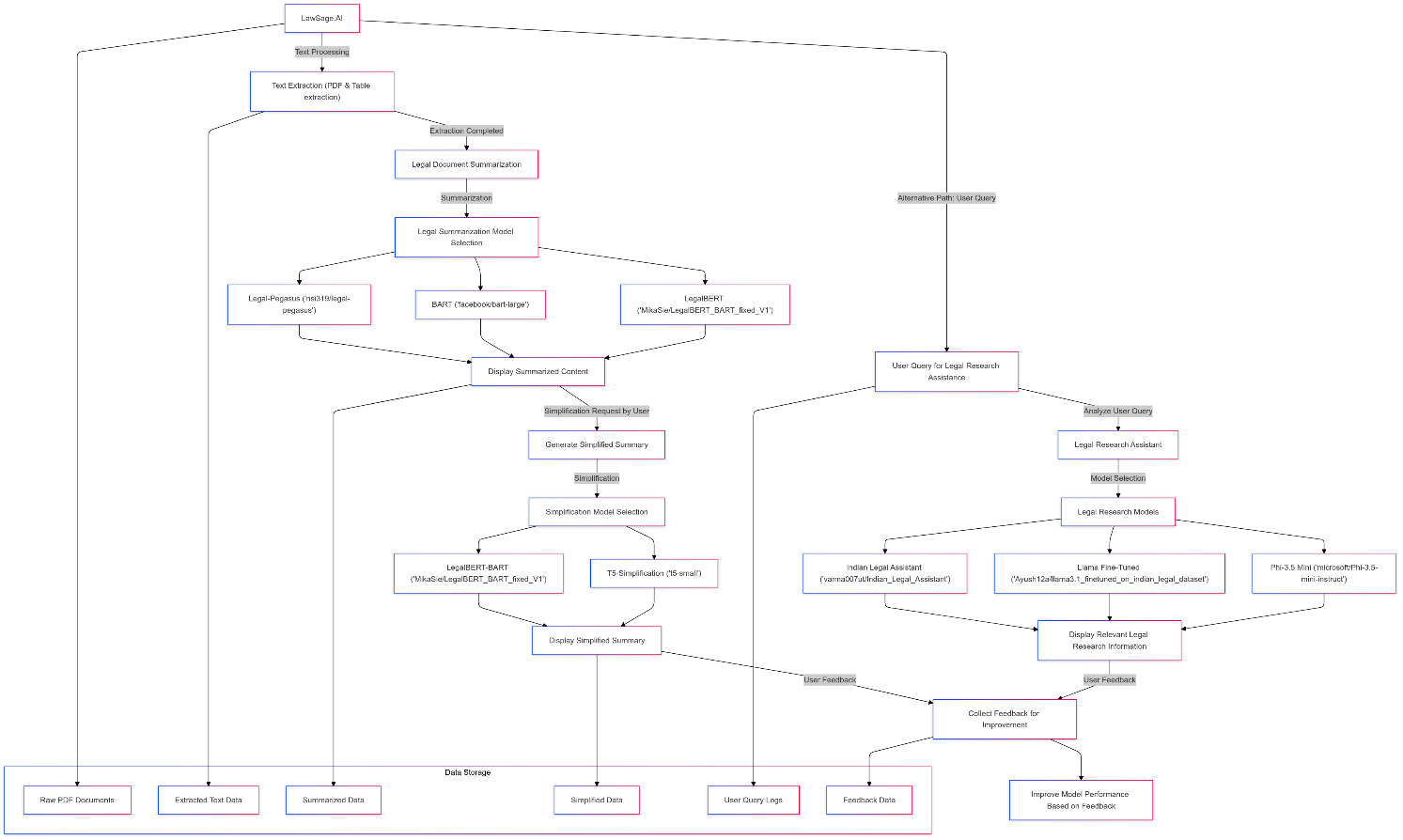
**Literature Review**

In this section, we examine existing models and methods used for LDSS in recent years. Table below provides an overview of selected studies that have contributed to the advancement of legal document summarization and simplification in recent years.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STUDY** | **MODEL** | **METHOD** | **EVALUATION METRICS** | **SCORES** |
| Parikh et al., 2021 | Weakly Supervised Sentence Extractor | Auto-labelling technique for identifying sentences with summary-worthy information and weakly supervised training on this auto-labelled data. | R1-F1 Score ,  R1-Precision,  R1-Recall | R1-F1 Score :  0.577  R1-Precision :  0.580  R1-Recall : 10.619 |
| Sharma & Singh, 2024 | InLegalBERT (optimized for summarization) | Using InLegalBERT for downstream summarization tasks; comparison with Legal Pegasus, T5 base, BART, BERT | ROUGE-1 F1, ROUGE-2 F1, ROUGE-L F1, Precision, Recall | ROUGE-1 F1: 0.4226, ROUGE-2 F1: 0.2604, ROUGE-L F1: 0.4023, Precision: 0.3022, Recall: 0.664 |
| Ghosh et al., 2022 | BART, PEGASUS | Normalizing Indian legal texts for summarization; BART for extractive, PEGASUS for abstractive summarization | ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-L (Precision, Recall, F-score) | ROUGE-1: 0.46/0.55/0.48, ROUGE-2: 0.28/0.36/0.31, ROUGE-3: 0.21/0.27/0.24, ROUGE-L: 0.35/0.45/0.39 |
| Sen et al., 2024 | Pre-trained GPT-2 | Greedy Search,  Beam Search,  Top-k Sampling,  Top-p Sampling,  Contrastive Searching,  Locally Typical Searching | Perplexity,  BLEU,  ROUGE,  METEOR,  TER | Not explicitly provided but showed the performance table of each method for each evaluation metrics. |
| Trivedi et al., 2023 | LED, BigBird (Long Document Transformer Models) | Abstractive Text Summarization | ROUGE-1, ROUGE-2, ROUGE-L, (F1/Recall) | ROUGE-1 : 15.69/ 9.5,  ROUGE-2: 6.02 3.5,  ROUGE-L :14.48/9 |
| Quevedo et al., 2023 | Various Legal NLP Models (Not specified) | Systematic Mapping Study (SMS) | F1 Score | F1 Score : 0.973 |
| Lai et al., 2023 | The paper does not specify a particular model used in the study. However, it mentions LLMs in the context of the legal field, which encompasses several models. | The paper is a comprehensive survey of legal LLMs, their applications, challenges, and future directions in the judicial industry. | The paper does not detail specific evaluation metrics | - |
| He et al., 2024 | JuriSim, JuriSim\_Lawformer | Dual Residual Cross-Attention Learning | Macro-F1 for charge and penalty prediction | 3.45% improvement in Macro-F1 for charge prediction, 3.05% improvement for term of penalty prediction |
| Perez-Castro et al., 2023 | GPT-2 (Generative Pre-trained Transformer 2) | Temperature | Comparison with state-of-the-art baselines (FIRE RISOT, TREC datasets) | Significant improvements over baselines on Bangla, Hindi, and English datasets |
| Jain et al., 2023 | BART (Bidirectional and Auto-Regressive Transformer) | Extract-Then-Assign (ETA): Builds an abstractive dataset by generating multiple extractive summaries (512 token length) for each document and assigning ground-truth summary sentences to them to generate new training samples for fine-tuning. | ROUGE, BERTScore | Improvement observed on FIRE and BillSum test sets in terms of ROUGE and BERTScore metrics, but exact percentage scores are not specified. |
| Xiao et al. (2021) | Lawformer | Utilizes a Longformer-based architecture to handle long-range dependencies in Chinese legal documents. | F1-score, Precision, and Recall. | high performance on these metrics compared to baseline models. |
| Zhong et al. (2020) | BERT and BERT-MS | Element extraction tasks like recognizing key legal details in cases involving divorce, labor, and loan disputes. | Micro-F1 (MiF) and Macro-F1 (MaF) | For BERT: MiF = 83.3, MaF = 69.6.  For BERT-MS: MiF = 84.9, MaF = 72.7. |
| Zhou et al. (2024) | LawGPT | Combines legal-oriented pre-training on a large corpus of Chinese legal documents with legal-specific supervised fine-tuning using a knowledge-driven instruction dataset. | Performance was benchmarked against the open-source LLaMA 7B model. | While specific numeric scores aren’t directly provided in the paper’s abstract, experimental results indicate that LawGPT outperformed LLaMA 7B on key Chinese legal tasks. |
| Zhang et al. (2024) | Not specific to a particular model but focuses on evaluating LLMs in the legal domain. | Proposes a novel evaluation methodology. | - | - |
| Dai et al. (2023) | LAiW (Legal AI Workbench) | It categorizes legal tasks into three difficulty levels based on legal practice logic: basic information retrieval, legal foundation inference, and complex legal application. This framework aligns with the thinking and reasoning process of legal professionals, using a structured reasoning approach as a guiding principle. | Human evaluation involves legal experts. | - |
| Satterfield et al. (2024) | Llama 3 (8 billion parameters) | Fine-tuning Llama 3 on legal texts, specifically case law data from Google Scholar, using a detailed approach involving model configuration adjustments and iterative training cycles to optimize performance on the LegalBench dataset. | Accuracy, Precision, Recall, and F1-score | Pre-Tuning: 80, 77, 70,72.  Post-Tuning: 93, 89, 83, 85 |
| Chun and Elkins (2024) | This study conducts an ethics-based audit on eight leading commercial and open-source LLMs. | The researchers designed ethical dilemmas that challenge the LLMs' adherence to or deviation from normative principles, testing each model's reasoning and decision-making. | - | - |
| Choi and Schwarcz (2023) | Empirical study investigates whether AI, specifically GPT-4, can replace human legal reasoning and improve  performance on law school exams | The study includes getting assessment results from students exposed to AI or not.The performance is evaluated on both MCQ's and complex questions. | - | - |
| Ghosh et al. (2024) | Explores the role of LLMs in Legal Text Analytics, specifically for the Indian legal system. | The paper presents a human focused combined Ai system merging LLM's with  human knowledge and highlighting the need to include human contributions for increased model efficacy and performance to be on par with legal professionals | - | - |
| Liga and Robaldo (2023) | Fine-tuning GPT-3 on a custom dataset | The model is trained to classify different legal rules within the document. | For evaluation, the authors used accuracy as the main metric to assess the model’s performance in classifying the legal rules accurately | - |

The survey indicates transformer-based models have taken a lead in LDSS by providing both extractive and abstractive methods. This can be selected on the basis of the legal document needed. For our another feature of legal assistant we may look for pre-trained LLMs which opensource available and then finetune them, according to our need.

**Propose Work**



Our work in developing a responsive web application where the user, who can be anyone, may be a student, legal professional, or a common person, can either summarize and simplify or get their legal query resolved. For frontend work, we used HTML/CSS and JavaScript. For the backend, the Flask framework was used. Python is the language used in backend logic and AI models. PyTorch and TensorFlow frameworks were used for model training and integration. We did empirical analysis of several models to check which could give us the best output. We also use Hugging Face models, which are open-source and have proven on several evaluation metrics. The specific models that were used in the project are shown above in the diagram.

**Experimental Results**

The results that we achieved are quite impressive. From the result of our analysis, we used Microsoft’s Phi-3.5-mini-instruct and Indian\_Legal\_Assistant: A LLaMA-based model for Indian Legal Text Generation for our legal research assistant. The Phi-3.5 model is a 3.82B parameter model that has an average 84.1 RULER score and an average 77 RepoQA score. The other assistant model is of the 8.03B parameter, which also has proven scores as it is a fine-tuned version of the LLaMA model. For our second feature of LDSS, we used the legal-pegasus model and LegalBert\_BART\_fixed\_v1. The legal-pegasus has decent scores as: rouge1 (57.39), rouge1-precision (62.97), rouge2 (26.85), rouge2-precision (28.42), rougeL (30.91), and rougeL-precision (33.22).

**Conclusion**

This web application aims to ease the legal queries and provide simplified summarization of lengthy text filled with jargon to the user. The application is ready for production deployment so that anyone across India can get the issues resolved. This work was made as part of an academic project for Intelligent Model Design using AI course subject. The developers are happy to share their work with the public. Till now, whatever is done, we consider it version one of the application, since as we were developing this we caught new ideas that could enhance the application. There are many future works that are needed to make this application more robust.

**Future Scope**

Our work holds a strong potential for performing best on the ground. In the future, we have planned to use more sophisticated models for either feature of LawSage.ai. We also aim to get financial benefit from the web application by using AdSense and enabling a premium feature option. The premium feature option will unlock more robust models and increase input token length. For now, the platform works only in the English language, but in the future it may have the capability to support multilingualism. Voice-Assisted Legal Queries can enhance accessibility for legal practitioners and people who are less educated and have low ability to read and understand.

**References**

1. Parikh, V., Mathur, V., Mehta, P., Mittal, N., & Majumder, P. (2021). LawSum: A weakly supervised approach for Indian Legal Document Summarization. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2110.01188
2. Sharma, S., & Singh, P. P. (2024). Domain-Specific Summarization: Optimizing InLegalBERT for Indian Judgment Reports. Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-3792484/v1
3. Ghosh, S., Dutta, M., & Das, T. (2022). Indian Legal Text Summarization: A Text Normalisation-based Approach. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2206.06238
4. Sen, J., Pandey, R., Waghela, H., Rakshit, S. (2024). Generative AI-Based Text Generation Methods Using Pre-Trained GPT-2. DOI: 10.13140/RG.2.2.20559.50082
5. Trivedi, P., Jain, D., Gite, S., Kotecha, K., Bhatt, A., & Naik, N. (2023). Indian Legal Corpus (ILC): A Dataset for A dataset summariz-ing Indian Legal Proceeding using Natural Language. Engineered Science. https://doi.org/10.30919/es1022
6. Quevedo, E., Cerny, T., Rodriguez, A., Rivas, P., Yero, J., Sooksatra, K., Zhakubayev, A., & Taibi, D. (2023). Legal Natural Language Processing from 2015-2022: A Comprehensive Systematic Mapping Study of Advances and Applications. IEEE Access, 1. https://doi.org/10.1109/access.2023.3333946
7. Lai, J., Gan, W., Wu, J., Qi, Z., & Philip S. Yu. (2023). Large Language Models in Law: A Survey. In Preprint. https://doi.org/10.13140/RG.2.2.15031.09124
8. He, C., Tan, T., Xue, S., & Tan, Y. (2024). Simulating judicial trial logic: Dual residual cross-attention learning for predicting legal judgment in long documents. Expert Systems With Applications, 125462. https://doi.org/10.1016/j.eswa.2024.125462
9. Perez-Castro, A., Martínez-Torres, & Toral, S. (2023). Efficiency of automatic text generators for online review content generation. Technological Forecasting and Social Change, 189, 122380. https://doi.org/10.1016/j.techfore.2023.122380
10. Jain, D., Borah, M. D., & Biswas, A. (2023). Summarization of Lengthy Legal Documents via Abstractive Dataset Building: An Extract-then-Assign Approach. Expert Systems With Applications, 237, 121571. https://doi.org/10.1016/j.eswa.2023.121571
11. Xiao, C., Hu, X., Liu, Z., Tu, C., & Sun, M. (2021). Lawformer: A pre-trained language model for Chinese legal long documents. AI Open, 2, 79–84. https://doi.org/10.1016/j.aiopen.2021.06.003
12. Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., & Sun, M. (2020). How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2004.12158
13. Zhou, Z., Shi, J., Song, P., Yang, X., Jin, Y., Guo, L., & Li, Y. (2024). LawGPT: A Chinese Legal Knowledge-Enhanced Large Language Model. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2406.04614
14. Zhang, R., Li, H., Wu, Y., Ai, Q., Liu, Y., Zhang, M., & Ma, S. (2024). Evaluation Ethics of LLMs in Legal Domain. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2403.11152
15. Dai, Y., Feng, D., Huang, J., Jia, H., Xie, Q., Zhang, Y., Han, W., Tian, W., & Wang, H. (2023). LAiW: A Chinese Legal Large Language Models Benchmark (A Technical Report). arXiv (Cornell University). https://doi.org/10.48550/arxiv.2310.05620
16. Satterfield, N., Holbrook, P., & Wilcox, T. (2024). Fine-tuning Llama with Case Law Data to Improve Legal Domain Performance. Cogni-Law Analytics. <https://doi.org/10.31219/osf.io/e6mjs>
17. Chun, J., & Elkins, K. (2024). Informed AI Regulation: Comparing the Ethical Frameworks of Leading LLM Chatbots Using an Ethics-Based Audit to Assess Moral Reasoning and Normative Values. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2402.01651
18. Choi, J. H., & Schwarcz, D. (2023). AI Assistance in Legal Analysis: An Empirical Study. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4539836
19. Ghosh, S., Verma, D., Ganesan, B., Bindal, P., Kumar, V., & Bhatnagar, V. (2024). Human Centered AI for Indian Legal Text Analytics. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2403.10944>
20. Liga, D., & Robaldo, L. (2023). Fine-tuning GPT-3 for legal rule classification. Computer Law & Security Review, 51, 105864. https://doi.org/10.1016/j.clsr.2023.105864