Enhancing Summarization of Legal Text Documents using Pre-trained Models.

Ashish Kasar
Department of Artificial Intelligence and Data Science
Vishwakarma Institute of Information Technology
Pune, India
ashish.22210968@viit.ac.in

Suyash Matade
Department of Artificial Intelligence and Data Science
Vishwakarma Institute of Information Technology
Pune, India
suyash.22210966@viit.ac.in

• Durvankur Rasal
Department of Artificial Intelligence and Data Science
Vishwakarma Institute of Information Technology
Pune, India
durvankur.22210306@viit.ac.in

Swapnil Shinde
Department of Artificial Intelligence and Data Science
Vishwakarma Institute of Information Technology
Pune, India
swapnil.shinde@viit.ac.in

***Abstract*—A very relevant challenge of summarizing legal text documents often exceeds the length, therefore being complex, was addressed using pre-trained models such as BART and PEGASUS. In the paper, the issue is discussed-the challenge that these pre-trained models face while being effective in general text summarization tasks, which becomes difficult in legal contexts due to specialized vocabulary and structure of legal documents. The authors fine-tune these models using the BillSum dataset-provided US Congressional bills as a proxy for court documents. Evaluation on ROUGE metrics reveals that BART outperforms PEGASUS in summation of content structures of legal texts to the greatest extent intended. The results suggest that the pre-trained models significantly improve the process of summarizing legal documents, both for professionals in law and general readers who would otherwise have to wade through an extensive amount of legal information.**

***Index Terms*—Legal text summarization, Natural Language Processing (NLP), Pre-trained language models, BART model, PEGASUS model, Abstractive summarization, Legal documents, BillSum dataset, Deep learning, ROUGE scores, Fine-tuning models, Model evaluation, Machine learning, Text tokenization, Transformer models, Legal domain-specific summarization.**

I. Introduction

In Natural Language Processing (NLP), [1] text summarisation is an essential activity that attempts to produce clear, cohesive summaries from lengthy text documents. The growing amount of legislation and case law material in the legal field presents a major difficulty for legal professionals who have to go through and extract pertinent information from long texts [2]. Hand summarisation is a tedious task since court case records have complex vocabulary and organisation. Automated summarisation is a solution to this problem that reduces the need for human effort while increasing [3] the accessibility of legal information. The development of deep learning and pre-trained language models has greatly enhanced the capacity to produce precise and fluid summaries. Cutting-edge models that use enormous volumes of data for pre-training, like as BART and PEGASUS, have demonstrated remarkable performance in a variety of text summarisation tasks. [4] Nonetheless, the specialised context and distinct lexicon found in legal writings make them domain-specific, posing issues that need for specialised methods for efficient summarisation. In this paper, we investigate the possible advantages of improving court document summarisation through the use of pre-trained models, namely BART and PEGASUS. [5] We use BillSum, a collection of US Congressional bills, in place of actual court records. We use the BillSum dataset to fine-tune these models with the goal of improving the legal domain's abstractive summarisation performance. This work's approach, analysis, and findings show how these trained models can help the public and legal experts alike by reducing the amount and complexity of legal texts.

II. Related Work

Legal document summarization has advanced significantly with the advent of Natural Language Processing (NLP) and Machine Learning (ML) approaches. Abstractive techniques have been the focus of current research on various tactics aimed at enhancing the quality and accuracy of summaries [6]. ArgLegalSumm recommends employing argument role labeling as one technique to improve abstractive summarizing. [7] The previously indicated methodology ensures that the argumentative framework of legal texts is retained in the resulting summaries, which is a crucial component in maintaining the credibility of legal reasoning [8]. An overview of the court documents A summary model designed specifically for legal information is produced by fusing machine learning and conventional NLP techniques with NLP and ML. Tokenization, cleaning, and other preparatory operations must be completed in advance when using this method [9]. Legal Documents in Hindi (HLDC) introduces a fresh corpus made especially

979-8-3315-2210-0/25/$31.00 ©2025 IEEE

to help in summarizing legal papers in Hindi. [10] This corpus was developed by the acquisition of a large number of Hindi-language legal books. To guarantee uniformity, which is essential for efficient summary, these texts underwent stringent text normalizing procedures. This research focuses on the application of a comprehensive text normalization technique because Indian legal texts usually comprise complex sentence structures and a wide variety of languages. Improving Legal Document The field of summarization studies the use of sophisticated natural language processing (NLP) models, especially pre-trained transformers, to enhance summary results. [11] Three steps make up the process: evaluation, model training, and data preparation. This work demonstrates how trained models such as BART and PEGASUS can produce summaries. An overview of the court records A model that makes use of hierarchical attention processes to enhance legal content summaries is presented in Using Hierarchical Attention Networks [12]. A Survey on Abstractive summarizing of Legal Documents provides a thorough overview of the state-of-the-art methods for abstractive summarizing in the legal profession. This research classifies the techniques based on the underlying algorithms, highlighting their respective benefits and drawbacks. Zhang publish Law-Pegasus, a specialized version of the Google Pegasus model, while Shukla explore the use of models such as BART and Legal-LED to document decisions in Indian and UK courts. , is a matter of discreet Filippova and Zhao tried to address the issue of Master's intelligent but this is still a concern, especially in law where truth matters. Although the abstraction process is promising in terms of slightly higher ROUGE and BLEU scores, illusions and inconsistencies limit their application without human interfere The lack of big data is a major problem. [13] Early research focused on legal models such as Letsum for Canadian decisions in Western countries such as Canada, the United Kingdom, and the United States Grishman and the Case Summarizer tool for the Australian legal system [14].

III. Methodology

The BillSum dataset is used to fine-tune pre-trained models, namely BART and PEGASUS, on a legal document summarizing job as part of the suggested methodology for improving the summary of legal case documents. The three primary phases of the process are Model Training, Model Evaluation, and Data Preparation. Furthermore, implementation details are given to guarantee clarity and reproducibility.

1. Data Preparation We used the publicly available BillSum dataset, which contains summaries of US Congressional bills, as a stand-in for court records. The selection of BillSum was motivated by the similarities between court documents and legislative bills, particularly in regard to their structured content and complex vocabulary. Both require precise summaries of lengthy and complex materials.We have used 100 training samples and 50 testing samples for fine tune the model. Data Preprocessing part includes following: • Shuffling: To minimize the biases and ensure representative distribution of samples, we used Shuffling. • Tokenization: By using Hugging Face Transformers library, We tokenized the text and summaries in dataset. We set a maximum length of 512 tokens for input texts and for summaries, 150 tokens. Also padding was applied to maintain uniform input sizes of batches. 2.Model Training: For this task we have chosen two complex pre-trained models: BART (facebook/bart-large-cnn): Known for its skill in extractive summarization, as well as other seq2seq tasks. Abstractive Summarization Expert PEGASUS (google/pegasus-large) The fine-tuning was executed with Hugging Face Trainer API to make model training simple. [2] Important setups comprised: Set the batch size to 1 per device in order for you to minimize the memory used. 16 steps of gradient: This is a method to simulate the larger batch size in the limited memory environment in order to improve training stability.

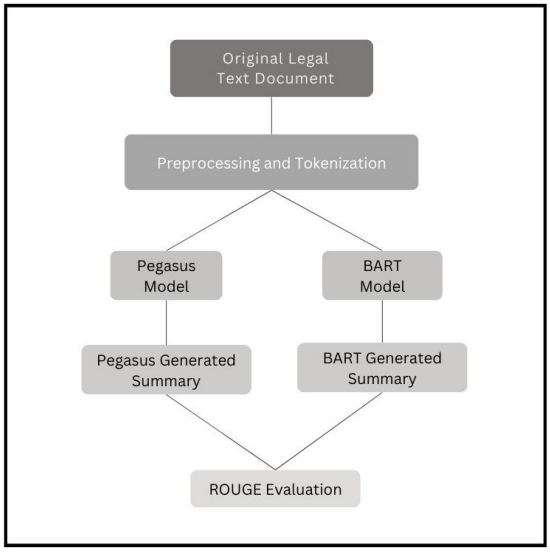

Fig. 1. Process Flow

This is why we have given two epochs since it has constraints. Check pointing and logging: Configured to save checkpoints of the model with some periodically interval and log progress. [1] Evaluation of the Model Using the ROUGE metric package, we assessed the performance of the model by concentrating on: ROUGE-1: Measures overlap of unigrams. ROUGE-2: Bigram overlap is evaluated by. ROUGE-L:The longest common subsequence is assessed using These metrics offer a thorough evaluation of the generated summaries' quality in relation to reference summaries.

Evaluation Procedure Importing Fine tuned Models: loaded both of the optimized PEGASUS and BART model in their respective directories. Generating Summaries: Using the models, summaries of 100 validation texts from the Billsum dataset were generated. The summarizing method was further configure with a maximum summary length of 60 tokens and four beams for beam search resulting in an higher quality final summary. Finding out ROUGE Scores The resulting summaries were compared against reference summaries using ROUGE-1, ROUGE-2, and ROUGE-L scores.

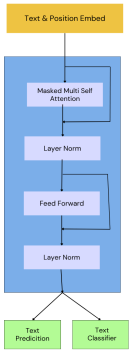

Fig. 2. BART Model Architecture

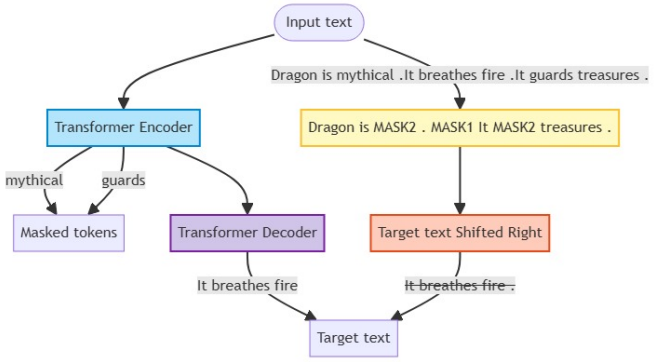

Fig. 3. PEGASUS Model Architecture

IV. Evalutation

The project was further designed to complete an evaluation phase in order to quantitatively assess the performance of newly fine-tuned BART and PEGASUS models when summarized legal documents. The primary evaluation tool used was the ROUGE metric package, a common method of comparing generated summaries with reference summaries. In the remainder of the post, we describe our evaluation process and results.

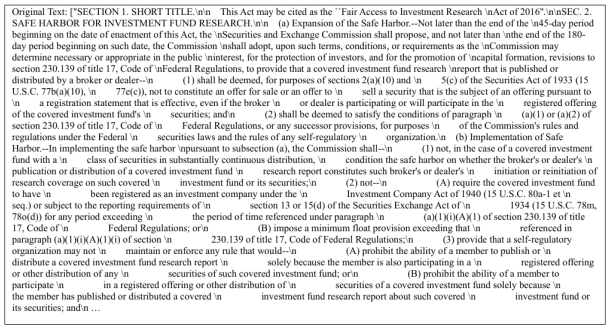
Original Text: ["SECTION 1. SHORT TITLE.\n\n This Act may be cited as the ''Fair Access to Investment Research \nAct of 2016".\n\nSEC. 2. SAFE HARBOR FOR INVESTMENT FUND RESEARCH.\n\n (a) Expansion of the Safe Harbor.--Not later than the end of the \n45-day period beginning on the date of enactment of this Act, the \nSecurities and Exchange Commission shall propose, and not later than \nthe end of the 180-day period beginning on such date, the Commission \nshall adopt, upon such terms, conditions, or requirements as the \nCommission may determine necessary or appropriate in the public \ninterest, for the protection of investors, and for the promotion of \ncapital formation, revisions to section 230.139 of title 17, Code of \nFederal Regulations, to provide that a covered investment fund research \nreport that is published or distributed by a broker or dealer--\n (1) shall be deemed, for purposes of sections 2(a)(10) and \n 5(c) of the Securities Act of 1933 (15 U.S.C. 77b(a)(10), \n 77e(c)), not to constitute an offer for sale or an offer to \n sell a security that is the subject of an offering pursuant to \n a registration statement that is effective, even if the broker \n or dealer is participating or will participate in the \n registered offering of the covered investment fund's \n securities; and\n (2) shall be deemed to satisfy the conditions of paragraph \n (a)(1) or (a)(2) of section 230.139 of title 17, Code of \n Federal Regulations, or any successor provisions, for purposes \n of the Commission's rules and regulations under the Federal \n securities laws and the rules of any self-regulatory \n organization.\n (b) Implementation of Safe Harbor.--In implementing the safe harbor \npursuant to subsection (a), the Commission shall--\n (1) not, in the case of a covered investment fund with a \n class of securities in substantially continuous distribution, \n condition the safe harbor on whether the broker's or dealer's \n publication or distribution of a covered investment fund \n research report constitutes such broker's or dealer's \n initiation or reinitiation of research coverage on such covered \n investment fund or its securities;\n (2) not--\n (A) require the covered investment fund to have \n been registered as an investment company under the \n Investment Company Act of 1940 (15 U.S.C. 80a-1 et \n seq.) or subject to the reporting requirements of \n section 13 or 15(d) of the Securities Exchange Act of \n 1934 (15 U.S.C. 78m, 78o(d)) for any period exceeding \n the period of time referenced under paragraph \n (a)(1)(i)(A)(1) of section 230.139 of title 17, Code of \n Federal Regulations; or\n (B) impose a minimum float provision exceeding that \n referenced in paragraph (a)(1)(i)(A)(1)(i) of section \n 230.139 of title 17, Code of Federal Regulations;\n (3) provide that a self-regulatory organization may not \n maintain or enforce any rule that would--\n (A) prohibit the ability of a member to publish or \n distribute a covered investment fund research report \n solely because the member is also participating in a \n registered offering or other distribution of any \n securities of such covered investment fund; or\n (B) prohibit the ability of a member to participate \n in a registered offering or other distribution of \n securities of a covered investment fund solely because \n the member has published or distributed a covered \n investment fund research report about such covered \n investment fund or its securities; and\n …

Fig. 4. legal doc input

The main evaluation metrics we used to select were the Recall-Oriented Understudy for Gisting Evaluation(ROUGE)

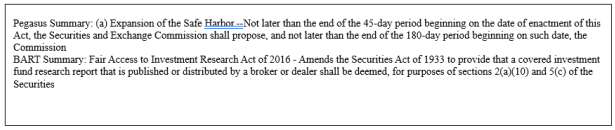
Pegasus Summary: (a) Expansion of the Safe Harbor.--Not later than the end of the 45-day period beginning on the date of enactment of this Act, the Securities and Exchange Commission shall propose, and not later than the end of the 180-day period beginning on such date, the Commission
BART Summary: Fair Access to Investment Research Act of 2016 - Amends the Securities Act of 1933 to provide that a covered investment fund research report that is published or distributed by a broker or dealer shall be deemed, for purposes of sections 2(a)(10) and 5(c) of the Securities

Fig. 5. legal doc output

TABLE I
Results and Performance

Evaluation Parameter's	BART Model	PEGASUS Model
ROUGE-1 Score	0.3634	0.2483
ROUGE-2 Score	0.2176	0.0967
ROUGE-L Score	0.2901	0.1703

metrics.

These are some of the other metrics along with KL divergence which are commonly used for quality metrics of text summarization models. They calculate the matches between generated summaries and reference summaries at multiple levels: The ROUGE-1 statistic measures the extent to which the produced summary and the reference summary share words, or unigrams. An rudimentary measure of content overlap is provided. The ROUGE-2 measure evaluates the degree of overlap between the generated and reference summaries in bigrams, which are groups of two words. The degree to which the produced summary properly portrays the links between words is indicated. ROUGE-L: The longer common subsequence (LCS) that divides the produced and reference summaries is the main focus of this metric. By taking word order and gaps into account, it offers a more flexible method of assessing summary quality. These three metrics together provide a comprehensive evaluation of the generated summaries, covering basic content accuracy (ROUGE-1), fluency and coherence (ROUGE-2), and structural similarity (ROUGE-L).

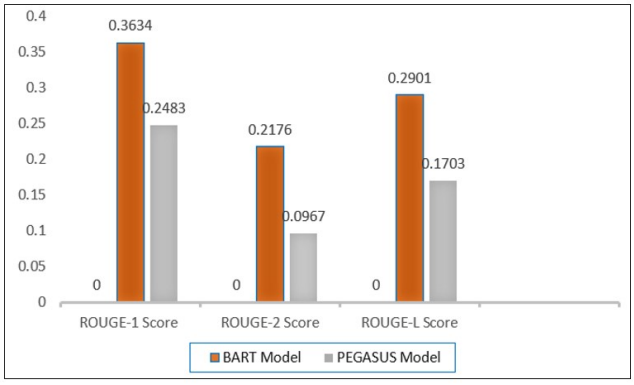

Fig. 6. Result Comparison

Performance Analysis of PEGASUS Abstractive summarization There may be many models to use for abstractive text summarization, like the PEGASUS model NLP which showed

quite good but not that much on this task. The ROUGE-1 score of 0.2483 demonstrates that PEGASUS learned some of the content material within the default summaries, but its ROUGE-2 and L scores are significantly a great deal decrease at zero.9037 and 0.l703 for output-generated phraseings needed better coherence and whole organization structure compare to generated phrases in the reference summary text. The complexity of the legal text can be an explanation for this result: the PEGASUS model may need a more specialized fine-tuning or even more training data for it to get used to those types of texts.

Attacking BART PerformanceAnalysis Overall, the BART model performed best according to all three ROUGE metrics (ROUGE-1: 0.3634, ROUGE-2: 0.2176, and ROUGE-L: 0.2901). The results show (1) the ability of BART to capture the main content and structure of legal texts. Higher ROUGE-2 and ROUGE-L scores, in particular, show that BART could produce better summaries in terms of both content and structure. That lines up well with BART's proven strengths in both extractive and abstractive summarization use-cases.

Comparative Analysis: Our results illustrate that common pre-training and fine-tuning methods are effective at summarizing legal documents, specifically with respect to the downstream task of predicting case outcomes, and we find BART to be a more capable model for this specific task when compared with XSum. Higher ROUGE scores do indicate BART produces summaries more closely aligned to the reference texts in terms of both content and presentation. Although still a strong model, PEGASUS may need to be optimized further when producing 200B tokens in this area. The evaluation results demonstrate that BART outperforms PEGASUS in summarizing legal documents from the BillSum dataset. Using ROUGE metrics cleanly evaluated every model, so any additional incrementals we could do to better the performance in specific areas would go on! This evaluation is a key element of our approach in which all models not only trained but it has been well tested and validated to generate summaries with high precision for legal document.

V. Conclusion

Overall our work has shown that we can gain a significant improvement for summarization tasks on legal material using state-of-the-art machine learning algorithms such as BART. Particularly given the utility of fine-tuning and reliability assessment scores, BART makes a stronger case as a better model for this specific task as compared to PEGASUS. However PEGASUS is used for abstractive summarization, it didn't do well due to the complexity of the text. This creates more accurate and efficient automated summarizing systems in the legal sector then allows information management and traceability.

References

[1] P. Trivedi, D. Jain, S. Gite, K. Kotecha, A. Bhatt, and N. Naik, "Indian legal corpus (ilc): A dataset for a dataset summarizing indian legal proceedings using natural language," *Engineered Science*, vol. 27, 2 2024.

[2] S. Ghosh, M. Dutta, and T. Das, "Indian legal text summarization: A text normalisation-based approach," 6 2022. [Online]. Available: http://arxiv.org/abs/2206.06238 http://dx.doi.org/10.1109/INDICON56171.2022.10039891

[3] Galgani, "Knowledge acquisition with multiple summarization techniques for legal text," 2013. [Online]. Available: http://hdl.handle.net/1959.4/52678inhttps://unsworks.unsw.edu.auon

[4] H. Zhao, "Overview of judicial text summarization method," *Highlights in Science, Engineering and Technology CSIC*, vol. 5, 2024.

[5] D. Khargharia, "Applications of text summarization," *International Journal of Advanced Research in Computer Science*, vol. 9, pp. 76–79, 6 2018. [Online]. Available: http://ijarcs.info/index.php/Ijarcs/article/view/6037/4916

[6] I. Benedetto, L. Cagliero, F. Tarasconi, G. Giacalone, and C. Bernini, "Benchmarking abstractive models for italian legal news summarization," vol. 379, pp. 311–316, 12 2023.

[7] D. Anand and R. Wagh, "Effective deep learning approaches for summarization of legal texts," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, pp. 2141–2150, 5 2022.

[8] D. Jain, D. Borah, and A. Biswas, "Summarization of indian legal judgement documents via ensembling of contextual embedding based mlp models," 2021. [Online]. Available: http://ceur-ws.org

[9] D. L. Freire, A. M. de Almeida, M. de S. Dias, A. Rivolli, F. S. Pereira, G. A. de Godoi, and A. C. de Carvalho, "Legalsum: Towards tool for evaluation for extractive summarization of brazilian lawsuits," vol. 933 LNNS, pp. 258–267, 2024.

[10] I. Jagirdar, S. Gandage, B. Waghmare, and I. K. Student, "Enhancing legal document summarization through nlp models: A comparative analysis of t5, pegasus, and bart approaches," vol. 12, pp. 2320–2882, 2024. [Online]. Available: www.ijcrt.org

[11] R. C. Kore, P. Ray, P. Lade, and A. Nerurkar, "Legal document summarization using nlp and ml techniques," *International Journal of Engineering and Computer Science*, vol. 9, pp. 25 039–25 046, 5 2020.

[12] I. Glaser, S. Moser, and F. Matthes, "Summarization of german court rulings," pp. 180–189, 2021.

[13] S. A. Salihu, A. Musa, F. E. Usman-Hamza, A. G. Akintola, A. O. Balogun, H. A. Mojeed, and G. B. Balogun, "Automatic summarization of legal documents using sumy," *FUW Trends in Science Technology Journal, www.ftstjournal.com e-ISSN*, vol. 8, pp. 307–315, 2408. [Online]. Available: https://github.com/miso-belica/Sumy

[14] H. Nguyen and J. Ding, "Keyword-based augmentation method to enhance abstractive summarization for legal documents," pp. 437–441, 6 2023.