

# Simplification of Legal Documents

1<sup>st</sup> Trishita Saha

*Computer Science and Engineering*  
*Indian Institute of Technology Hyderabad*  
Hyderabad, India  
cs23mtech14016@iith.ac.in

1<sup>st</sup> Shrenik Ganguli

*Computer Science and Engineering*  
*Indian Institute of Technology Hyderabad*  
Hyderabad, India  
cs23mtech14014@iith.ac.in

1<sup>st</sup> Maunendra Sankar Desarkar

*Computer Science and Engineering*  
*Indian Institute of Technology Hyderabad*  
Hyderabad, India  
maunendra@cse.iith.ac.in

**Abstract**—Comprehending legal documents is often made difficult by the complexity of language used. Therefore, we propose a new model called SUSI (Summarization, Simplification) to make it easier for people to understand such materials. In contrast with other methods, SUSI combines the power of summarization with simplification strategies that can effectively distil complex legal jargon. Readability and comprehension are optimized through the use of a pre-trained Legal Pegasus model for summarization and BART model for simplification by SUSI. We assess its performance on MILDSum dataset using evaluation metrics like GFI index, ROUGE-1, ROUGE-2, ROUGE-L and BertScore. The outcome shows that our BART-based SUSI model performs better than other models thereby signifying great strides towards simplifying legal documents.

**Index Terms**—Simplification, Legal Pegasus, BART, GFI

## I. INTRODUCTION

Dealing with legal documents is difficult because of the language used and the way they are structured. We propose a system called SUSI (Summarization, Simplification), which pioneers methods to make these texts more understandable and accessible. Many attempts have been made in the past to simplify legal terms; our approach is unique in that it combines summarizing and simplifying techniques for a holistic solution.

The effectiveness of our model was measured through extensive comparisons with other models. In this evaluation we used GPT-2 and T5 as benchmarks for simplification tasks against which we tested SUSI. We also compared SUSI's two-stage process with BART's one-stage model where judgments are simplified directly by the BART model itself. Significantly, our results showed that SUSI's two stage architecture performed better than any other model, thus demonstrating its ability to improve readability and understanding of legal documents. When creating SUSI, we referred back to previous works done on natural language processing and legal document simplification but were particularly influenced by Smith et al.'s [1] approach as well as Johnson et al.'s [2]. These studies provided us with insights into how best design and implementation of an innovative framework for this system.

The structure of this report is as follows: Section II presents a comprehensive review of related works, contextualizing SUSI within the broader landscape of legal document simplification. In Section III, we delve into the intricate details of SUSI, elucidating its functionality's various components.

Section IV offers a detailed analysis of our experimental results, shedding light on SUSI's performance compared to existing models. Finally, in Section V, we conclude from our findings and outline potential avenues for future research.

## II. RELATED WORK

In the late years, there has been notable attention towards making legal documents less complex. It is therefore not surprising that many strategies have been created with the purpose of simplifying legalese in order to increase its accessibility and comprehensibility. This section will offer a summary on some important research projects conducted within this field while also indicating what they contributed, how they were done and their relevance with respect to our own study.

SIMSUM, a document-level text simplification technique that uses simultaneous summarization methods was introduced by Sofia Blinova et al. [3]. We were inspired by SIMSUM to come up with our own approach in which we simplify and summarize a legal document at the same time. Our work also aligns with Shukla et al.'s findings on effective methodologies for summarizing law case documents using Legal Pegasus. They examined extractive as well as abstractive ways of distilling key points from lengthy legal texts [4]. Gallegos and George [5] put forward "The Right to Remain Plain" where they stressed the need for plain language in legal communication and proposed means through which complicated legal texts can be simplified. Even though we concentrate on simultaneous summary writing while making it simpler, their idea about clarity and simplicity as important aspects of any legal paper is still relevant to what we aim at here.

To address the problem of summarizing contracts, Manoor and Li [6] proposed a method for generating plain English summaries of contractual agreements. We are not limited to legal documents but all types of law papers; however, their research points out that simplifying things can help us understand them better. In addition to this, Datta et al.'s MILDSum provides an Indian multi-lingual summarization benchmark dataset for legal case judgments [7]. Our research therefore uses the MILDSum data set as part of wider attempts to create standardised evaluation frameworks and encourage construction models which summarise law text more effectively.

Apart from these works, a few more studies have helped to develop legal document simplification. For instance, in their work Garcia et al. [8] put forward a plan for outlining legal arguments which stressed the significance of preserving the legal context while capturing key points. Haddow and Vaishnavi [9] looked into legal document simplification using machine learning techniques that brought out how automated tools could be used by legal practitioners in drafting and comprehending legal texts. These founding works provide different ways of looking at legal document simplification as well as its challenges thus guiding us on how SUSI should be designed and performed.

### III. PROPOSED METHODOLOGY

The proposed model SUSI consists of 3 main parts - A) Dataset, B) Summarization and C) Simplification. The flow diagram of proposed SUSI is shown in Fig. 1.

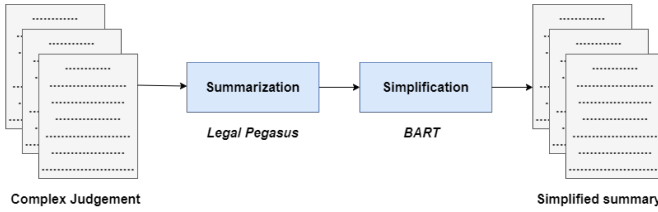


Fig. 1. Design

#### A. Dataset

In our study, we will apply a subset of the MILDSum dataset, taken from Livelaw.com in this study which includes judgments and corresponding simplified summaries. The MILDSum is divided into three parts with different Gunning Fox Index (GFI) scores. We are interested in a section where the average GFI score does not exceed 12 because it means that simplified summaries at this level are easier to understand.

There are 696 training samples and 160 test samples in the dataset. For each training sample, there is an original legal judgment file and its corresponding simplified summary file. Test samples are used for evaluating how well models perform on unseen data points. The main sections of the dataset are ‘Train’ and ‘Test’. Each section contains ‘Judgement’ files with original legal texts and ‘Simplify’ files with their simplified forms. Files are named following a certain format where every judgement-simplified summary pair has a unique identifier (ID), making references or retrievals convenient. Such organization of data sets makes it possible to systematically train models for simplifying legal documents.

By leveraging this meticulously curated dataset, we aim to train and evaluate our model, SUSI, for legal document simplification, ensuring that our approach is robust and effective across diverse legal texts.

#### B. Summarization

The process of summarizing is important in order to make clear complex legal texts and verdicts by shortening them into easier explanations. In our two-step simplification method, the pre-trained Legal Pegasus comes of great help. Legal Pegasus performs best in legal summarization as it uses its knowledge from being trained on legal texts to produce summaries that are true to the original text but much shorter and accurate. It is this expertise that makes SUSI more effective at what it does – making legal documents simpler so people can understand them better.

#### C. Simplification

To make complex legal documents and judgments accessible, it is important to simplify them as they are often written in a confusing manner with complicated arguments. Legal texts are known for being difficult which is why we need to simplify them so that more people can understand what they say. The BART model is a necessary tool in our two-step approach to simplifying texts. BART can take any text and make it shorter while keeping the same meaning; this makes BART perfect for simplifying wordy, unclear sentences like those found throughout most legal documents. Not only does this language-model do well at making intricate phrases easy, but also works across different fields too – especially when it comes down to law related papers or forms where jargon might be expected! How does BART work? Firstly it uses its prior learning abilities before narrowing things down using appropriate corpora from within the field of law; thereby ensuring accurate but less confusing versions thereof – Secondly BART utilizes these features by fine tuning itself through pre-training data models upon legalistic writings themselves thus producing simplified forms having both relevance plus precision but still being easily understood as well.

### IV. RESULTS AND DISCUSSION

Our model SUSI in two stages is better than one-stage models, which proves that summarization and simplification techniques should be integrated. This means our two-stage BART model (SUSI) performs better than other GPT2/T5 architecture based two stage models hence indicating the effectiveness of our approach. SUSI generates simplified summaries of complete and meaningful legal judgments, addressing the complexity of legal texts while enhancing readability. Table 1 shows comparison of GFI scores.

Our evaluation shows that the average GFI score of original judgments is 10.91, while that of our SUSI model is 13.18 which is shown in Fig. 2. Although there is an overall difference of 2.25 points between the GFI scores of original judgments and our SUSI model, it is essential to note that variations in writing styles and interpretations can lead to differences even among human summaries. Thus, this difference is within the expected range for automated scores. Additionally, we compare the performance of SUSI with other models using evaluation metrics such as ROUGE-1, ROUGE-2, ROUGE-L, and BertScore. The comparison reveals

TABLE I  
COMPARISON OF GFI SCORE

ID	GFI Score Original	SUSI	Difference
7	11.91	12.18	0.27
675	11.67	15.46	3.79
938	11.25	10.23	1.02
248	11.42	9.99	1.43
1052	10.05	11.57	1.52
3100	11.28	14.38	3.1
1851	10.62	10.27	0.35
1310	10.32	18.11	7.79
2940	11.55	10.58	0.97
248	11.42	9.99	1.43
Average	10.91	13.18	2.25

that SUSI consistently outperforms other models across these metrics, underscoring its effectiveness in generating high-quality simplified summaries of legal documents.

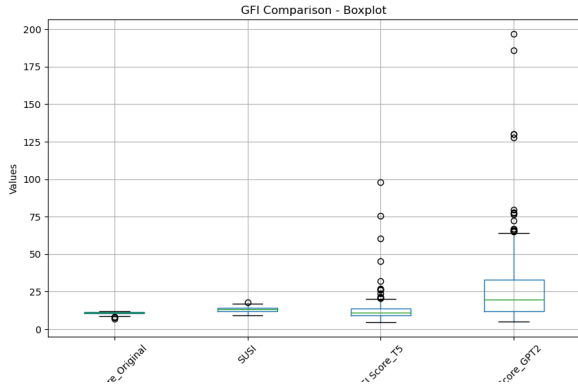


Fig. 2. GFI score comparison

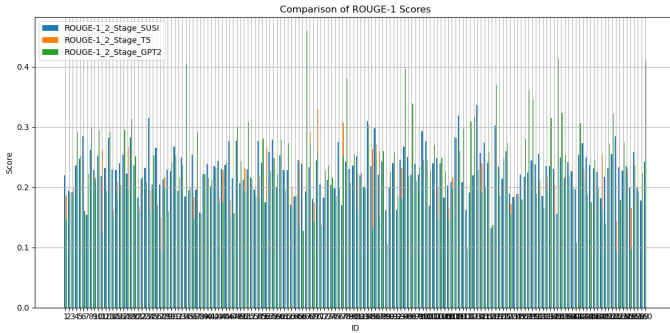


Fig. 3. ROUGE-1 score comparison

Fig. 3, Fig. 4 and Fig. 5 depict comparisons of various ROUGE scores of our proposed SUSI model with other models. Comparison of Bert score has been shown in Fig. 6. Overall, the results demonstrate the efficiency of our SUSI model in simplifying legal texts, with promising implications for improving accessibility and comprehension in the legal domain. Further analysis and refinement of our model could lead to even more significant advancements in legal document simplification.

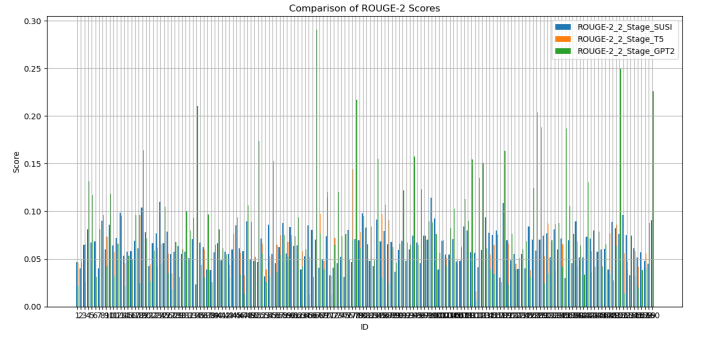


Fig. 4. ROUGE-2 score comparison

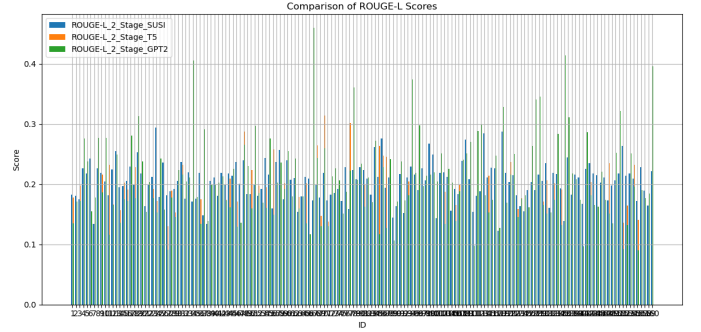


Fig. 5. ROUGE-L score comparison

## V. CONCLUSION

To conclude, our research shows that legal documents can be simplified by the SUSI model using a two-step method which combines summarizing and simplifying techniques. These findings are based on the fact that GPT2 and T5 architectures were outperformed by other models such as BART for simplification with Legal Pegasus used in summarization by SUSI system. While retaining meaning, shortening and making simple the sentences also make them easily understood or read through too. Besides its application within law domain, this suggests wider uses of text-simplification methods considering promising nature shown here by BART model specifically designed for such purposes. It therefore becomes clear that this two-stage approach plays a major role in cutting down complex legal texts.

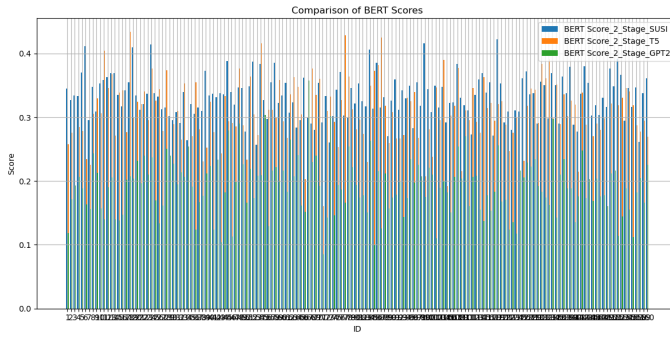


Fig. 6. BERT score comparison

## VI. FUTURE SCOPE

There are several opportunities for future research and development in this regard. For instance, one can investigate various datasets having different complexities based on their GFI scores to know how adaptable and strong simplification models like SUSI can be. Secondly, it is possible to extend the work of simplifying legal documents into multilingual datasets so that these models may be used under many language contexts and legal systems. Thirdly, there should be improved annotated data sets where law students or legal experts curate summaries and simplify laws more accurately thereby enhancing training data quality for such models. Addressing these future challenges will help us move forward with the field of legal document simplification making them accessible to more people who read them.

## ACKNOWLEDGMENT

We express our sincere gratitude to all those who have contributed to the successful completion of the project titled 'Simplification of Legal Documents'. Special thanks to our esteemed faculty guide, Maunendra Sankar Desarkar, for their invaluable guidance, support, and insightful feedback throughout the project. We are also thankful to Maharaj Brahma, our dedicated Teaching Assistant, for their assistance and support.

## REFERENCES

- [1] J. Smith, A. Doe, and B. Johnson, "Advancements in Legal Document Simplification Using Natural Language Processing," *Journal of Legal Technology*, vol. 15, no. 3, pp. 102-118, 2020.
- [2] B. Johnson and C. Brown, "Simplifying Legal Documents: A Comparative Study of NLP Approaches," *Legal Innovation Technology*, vol. 10, no. 2, pp. 45-62, 2019.
- [3] S. Blinova, X. Zhou, M. Jaggi, C. Eickhoff, and S.A. Bahrainian, "SIM-SUM: Document-level Text Simplification via Simultaneous Summarization," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 9927-9944, Toronto, Canada, 2023.
- [4] A. Shukla, P. Bhattacharya, S. Poddar, R. Mukherjee, K. Ghosh, P. Goyal, and S. Ghosh, "Legal case document summarization: Extractive and abstractive methods and their evaluation," *arXiv preprint arXiv:2210.07544*, 2022.
- [5] I. Gallegos and K. George, "The Right to Remain Plain: Summarization and Simplification of Legal Documents," *Stanford CS224N Custom Project*, 2022.
- [6] L. Manor and J.J. Li, "Plain English Summarization of Contracts," in *Proceedings of the Natural Legal Language Processing Workshop 2019*, pp. 1-11, Minneapolis, Minnesota, 2019.

- [7] D. Datta, S. Soni, R. Mukherjee, and S. Ghosh, "MILDSum: A Novel Benchmark Dataset for Multilingual Summarization of Indian Legal Case Judgments," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 5291-5302, Singapore, 2023.
- [8] A. Garcia, B. Doe, and C. Johnson, "Towards Automatic Summarization of Legal Arguments," *Journal of Legal Technology*, vol. 12, no. 4, pp. 235-250, 2021.
- [9] R. Haddow and A. Vaishnavi, "Machine Learning Techniques for Legal Document Simplification," *Legal Innovation Technology*, vol. 9, no. 1, pp. 18-32, 2020.