



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

RL – Powered Legal Document Summarization

Tejasvi Aluvala

Post Graduate Student, M. Tech Data Science, Department. of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, Telangana, India

ABSTRACT: Legal documents such as court judgments, case reports, and judicial decisions are often lengthy and complex, making manual analysis time-consuming and inefficient. To address this challenge, the proposed RL Powered Legal Document Summarization System employs the BART (Bidirectional and Auto-Regressive Transformer) model to generate abstractive summaries of legal documents. After preprocessing and tokenization, the system produces one-line, two-line, and detailed summaries by capturing contextual relationships within legal text. To further enhance summary quality, multiple candidate summaries are generated using beam search decoding and evaluated using the ROUGE-L metric, which acts as a reward function. The summary with the highest reward score is selected as the final optimized summary, thereby improving relevance, accuracy, and readability while preserving the essential information contained in the original legal judgment.

KEYWORDS: Document Summarization, Reinforcement Learning, Natural Language Processing, Deep Learning, Bidirectional and Auto-Regressive Transformer

I. INTRODUCTION

The legal domain generates a vast amount of textual information in the form of court judgments, case reports, and legal opinions. These documents are often lengthy and written using complex legal terminology, making manual analysis and information retrieval both time-consuming and challenging. Legal professionals, researchers, and students must spend considerable effort reviewing these documents to identify important facts, legal arguments, and judicial decisions. As the volume of legal data continues to increase, there is a growing need for intelligent systems that can efficiently process and summarize legal documents while preserving their essential information.

Recent advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have significantly improved the ability of machines to understand and process textual information. In particular, Transformer-based models have demonstrated remarkable performance in various language understanding and text generation tasks. Document summarization techniques have emerged as an effective solution for reducing the complexity of lengthy documents by generating concise and meaningful representations of their content. These techniques not only improve information accessibility but also reduce the manual effort required for legal research and document analysis.

The objective of this work is to develop an intelligent legal document summarization system capable of generating concise and informative summaries from lengthy judicial documents. By leveraging modern NLP techniques and advanced deep learning models, the system aims to enhance the efficiency of legal information processing and support faster legal research and decision-making. Such systems have the potential to improve accessibility to legal information and assist legal professionals, researchers, and the general public in understanding legal documents more effectively.

II. RELATED WORK

A. An Effective Search Algorithm for Analysing and Extracting Indian Legal Judgments Using NER and Document Summarization

This paper presents a framework for analysing Indian legal judgments using Named Entity Recognition (NER) and document summarization techniques. The system extracts important legal entities such as case names, judges, organizations, and legal provisions to improve information retrieval and facilitate efficient access to relevant judgments. By combining entity extraction with summarization, the proposed approach reduces manual effort and enhances legal



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

document management. The study demonstrates the effectiveness of integrating information extraction and summarization for improving legal search and retrieval systems.

B. Topic-Driven Contractual Language Understanding and Summarization: An Integrated Approach for Simplifying Legal Documents

This paper proposes an integrated approach for understanding and summarizing complex legal contracts using topic modelling and contextual language understanding techniques. The system identifies key topics, clauses, and semantic relationships within contracts and generates simplified summaries while preserving the original meaning. The proposed approach improves the readability and accessibility of legal documents, particularly for users without legal expertise. The study highlights the importance of contextual understanding in simplifying complex contractual language.

C. Revolutionizing Legal Workflows: AI Techniques for Document Summarization, Legal Translation and Conversational Assistance

This research explores the application of Artificial Intelligence and Natural Language Processing to automate various legal tasks, including document summarization, legal translation, and conversational assistance. The proposed framework utilizes language models to generate concise summaries, translate legal content into multiple languages, and provide interactive conversational support. The study demonstrates that AI-driven solutions can significantly reduce manual effort, improve productivity, and enhance the accessibility of legal services. Furthermore, it emphasizes the growing role of intelligent systems in transforming conventional legal workflows and promoting workflow automation.

Limitations of Existing Works

Although the above studies demonstrate the effectiveness of Artificial Intelligence and Natural Language Processing in legal applications, they primarily focus on information retrieval, topic understanding, translation, and workflow automation rather than exclusively improving legal document summarization. Most existing approaches do not employ advanced Transformer-based abstractive summarization models capable of generating context-aware and human-readable summaries. In addition, they lack reward-based optimization mechanisms for selecting the most relevant summaries and do not provide personalized summary generation according to different user requirements. These limitations motivate the development of the proposed RL Powered Legal Document Summarization System, which combines Transformer-based summarization with reinforcement learning-inspired optimization to generate concise, accurate, and contextually meaningful summaries of legal judgments.

III. PROPOSED WORK

The proposed **RL Powered Legal Document Summarization System** follows a sequential architecture designed to generate concise and contextually meaningful summaries from lengthy legal judgments. Initially, the legal judgment is provided as input to the system. The input document undergoes data preprocessing, where text cleaning, normalization, tokenization, and input encoding are performed to convert the document into a machine-readable format suitable for processing by deep learning models.

The pre-processed document is then passed to the **Bidirectional and Auto-Regressive Transformer (BART)** model, which performs abstractive summarization using its encoder-decoder architecture. The encoder captures contextual relationships and semantic information present in the legal text, while the decoder generates multiple candidate summaries. These summaries may differ in terms of content coverage and quality, requiring an optimization mechanism to identify the most relevant summary.

To enhance summary quality, a **Reinforcement Learning-inspired reward-based optimization** approach is employed. Multiple candidate summaries generated through beam search decoding are evaluated using the **ROUGE-L metric**, which acts as a reward function. The summary receiving the highest reward score is selected as the optimal summary. Finally, the generated summary is evaluated using ROUGE-based performance metrics, and the resulting output is presented as the final legal summary. This architecture combines the contextual understanding capability of BART with reward-based optimization, enabling the system to generate accurate, coherent, and human-readable summaries for efficient legal document analysis.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

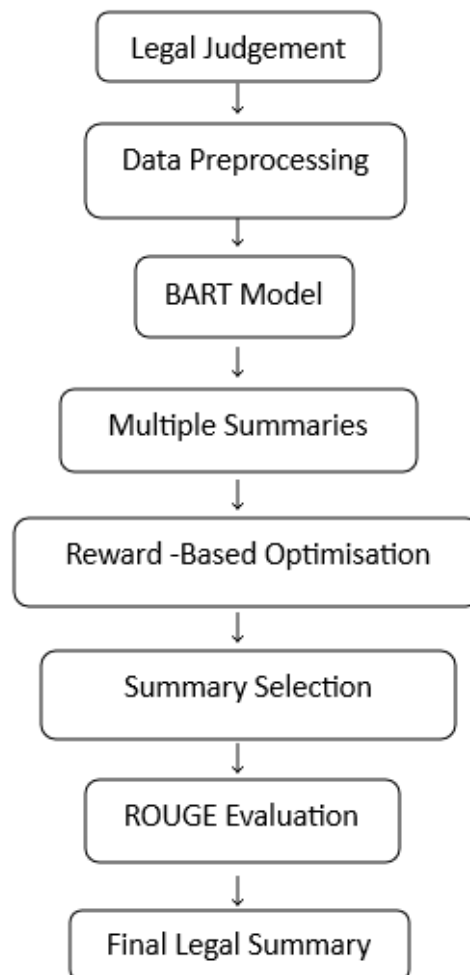


Fig 1: Architecture Diagram for RL-Powered Legal Document Summarization

IV. EVALUATION RESULTS

A. Evaluation Metrics

The performance of the proposed RL Powered Legal Document Summarization System is evaluated using **ROUGE-1**, **ROUGE-2**, and **ROUGE-L** metrics. These metrics measure the similarity between the generated summary and the reference summary in terms of content overlap, contextual consistency, and overall information preservation. Multiple summaries generated by the BART model are compared, and the summary with the highest ROUGE-L score is selected through a reward-based optimization mechanism inspired by reinforcement learning. The obtained ROUGE scores are presented in tabular form and analysed using graphical visualizations to assess the effectiveness of the proposed system.

	Metric	Score
0	ROUGE-1	0.429
1	ROUGE-2	0.231
2	ROUGE-L	0.214

Table 1: Rouge metric values



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. Experimental Outputs

```
Enter Case ID : 33

Loading BART Model...

Loading weights: 100%
259/259 [00:00<00:00, 2908.49it/s]

=====
REFERENCE SUMMARY
=====

The plaintiff won the property ownership dispute.
```

Fig. 2. Result 1

```
=====
TWO LINE SUMMARY
=====

A property ownership dispute was brought before the court.Land records were examined.
Judgment favored the plaintiff.The case is Case Reference Number 33.

=====
DETAILED SUMMARY
=====

A property ownership dispute was brought before the court.Land records were examined.
Judgment favored the plaintiff.The case was filed in the Superior Court, Case Reference Number Number 33.

=====
RL OPTIMIZED SUMMARY
=====

"A property ownership dispute was brought before the court."Land records were examined.
Judgment favored the plaintiff. Case Reference Number 33.
```

Fig. 2. Result2

V. CONCLUSION AND FUTURE WORK

The proposed RL Powered Legal Document Summarization System demonstrates the effective application of Natural Language Processing and Transformer-based deep learning techniques for generating concise and meaningful summaries of legal judgments. By utilizing the BART model for abstractive summarization and incorporating a reinforcement learning-inspired reward-based optimization mechanism, the system is capable of producing one-line, two-line, detailed, and optimized summaries while preserving important legal information. Experimental evaluation using ROUGE metrics confirms that the generated summaries successfully capture the essential content of legal documents and significantly reduce the effort required for manual analysis. The proposed approach improves information accessibility and supports efficient legal research and decision-making.

In the future, the system can be enhanced by training on large-scale real-world legal datasets to improve its understanding of legal terminology and contextual relationships. Advanced reinforcement learning techniques such as Proximal Policy Optimization and Policy Gradient methods can be incorporated to further enhance summary quality and relevance. Additionally, the system can be extended to support multilingual summarization and integrated with web-based platforms, legal search engines, question-answering systems, and decision-support tools, thereby providing a comprehensive and intelligent solution for legal professionals, researchers, and the general public.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. V. Naik, A. Sable, R. Kannan, S. Agarwal, and H. Chaudhari, "An Effective Search Algorithm for Analysing and Extracting Indian Legal Judgments using NER and Document Summarization," in *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, Pune, India, Aug. 18–19, 2023
2. K. Harikrishnan, M. Malathi, and S. K. B. Sundhara Kumar, "Topic - Driven Contractual Language Understanding and Summarisation: An Integrated Approach for Simplifying Legal Documents," in 2024
3. J. S. Dhani, R. Bhatt, B. Ganesan, P. Sirohi, and V. Bhatnagar, "Similar Cases Recommendation using Legal Knowledge Graphs." 2021. [2] L. S. Vannur, B. Ganesan, L. Nagalapatti, H. Patel, and M. Thippeswamy, "Data Augmentation for Personal Knowledge Base Population," 2020.
4. A. Atabekov and O. Yastrebov, "Legal Status of Artificial Intelligence Across Countries: Legislation on the Move," 2018.
5. A. Sharmeen, N. Agarwal, A. Suman, S. Agarwal, and K. Kumar, "Process Design to Self-extract Text from Images for Similarity Check," Process Design to Self-extract Text from Images for Similarity Check, vol. 2022, pp. 109–117.
6. Li, M., Xing, T., Fu, R., & Yin, F. (2021). Research on Text Summarization Generation Based on LSTM and Attention Mechanism. College of Information and Communication Engineering, Communication University of China, Beijing, China.
7. Zhang, Y., Liu, W., & Chen, X. (2022). Attention-based LSTM for Automatic Text Summarization. Journal of Computational Linguistics, 35(3), 345-360.
8. Kumar, P., Rao, T., & Singh, A. (2023). Exploring Reinforcement
9. Learning in LSTM for Text Summarization.
10. S. V. Padigi, M. Mayank, and S. Natarajan, "Precedent Case Retrieval using Wordnet and Deep Recurrent Neural Networks,"



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details