Automatic Summarization of Legal Documents

1st Aayush Goyal
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India

2nd Arnav Firke
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India

3rd Bhairu Gavhane
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India

4th Shrinidhi Kulkarni
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India

5th Pranjali Joshi
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India

***Abstract*—Automatic summarization of legal documents is a rapidly evolving field that leverages natural language processing (NLP) techniques to extract key information from lengthy and complex legal documents. Legal documents are often lengthy, complex, and filled with jargon that can be challenging to understand. Traditional methods of summarization require manual extraction of key information, which is time-consuming and prone to human error. There is a need for an automated system that can efficiently and accurately summarize legal documents, providing clear and concise information that highlights the key terms, obligations, rights, and conditions stipulated within the document.**

***Index Terms*—Text Processing, Natural Language Processing, Natural Language Generation, Summarization, Evaluation Methodologies, Information Extraction**

I. Introduction

Text summarization involves creating a coherent and succinct summary of lengthy documents, offering a brief context that is especially beneficial in specialized fields like law. In countries that adhere to the Common Law System, such as India, legal documents can be lengthy and filled with complex terminology, making manual summarization a labor-intensive task. With over 47 million cases pending in various courts across India, the potential for ATS to streamline the review of legal texts is significant.

Advancements in Artificial Intelligence (AI) and Machine Learning (ML) have facilitated the automation of text summarization, reducing the time and effort required for manual drafting. While there have been various techniques and tools developed for legal text summarization in countries like the UK, Canada, and Australia, the unique structure and terminology of Indian legal documents necessitate the creation of tailored solutions. To achieve this, a substantial dataset specific to Indian legal documents is essential, yet such datasets remain largely unavailable, hindering the development of effective domain-specific summarization models.

ATS systems are designed to accomplish objectives like as extracting the most important and relevant information from a document, generating summaries that are much shorter than the original content, etc.This includes various methods for generating summary which is discussed below.

Methods for Generating Summaries

There are three primary methods for generating summaries:

1) **Extractive Approach:** In this approach, important sentences from a document are picked and combined to generate a final summary. Major steps in an extractive approach include:
   a) Document pre-processing
   b) Create a provisional representation of the document
   c) Score sentences according to their retrieval value
   d) Select the sentences with the highest scores.
2) **Abstractive Approach:** This strategy seeks a much deeper comprehension of the document. Instead of selecting meaningful sentences directly, it generates new sentences that convey the same information using natural language processing algorithms. Important steps in an abstractive approach include:
   a) Preprocessing the document
   b) Making an intermediate representation of the document
   c) Generating new sentences based on information retrieval (IR).
3) **Hybrid Approach:** This approach combines both the abstractive and the extractive approaches to generate the summary.

II. Evaluation metrics

Automatic text summarization methods are assessed using various performance evaluation metrics. This section will explore these metrics in detail.

A. ROUGE (RECALL-ORIENTED UNDERSTUDY OF GISTING EVALUATION)

It is the most popular evaluation metric used in the field of text summarization. ROUGE has four types:

1) **ROUGE-N:** The ROUGE-N metric is commonly used to evaluate the overlap of n-grams between a candidate summary and a reference summary. For ROUGE-N, the general formula is:

$$\text{ROUGE-N} = \frac{\sum_{\text{gram}_n \in \text{Reference Summaries}} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{\text{gram}_n \in \text{Reference Summaries}} \text{Count}(\text{gram}_n)}$$

2) **ROUGE-L:** Here, L stands for longest common sub string. A sentence is represented as a set of words. The longer theLCSbetweenoursummaryandmanualsum mary sentences, the better the quality of the summary.
3) **ROUGE-W:** Here,WstandsforweightedLCS.Ittries the limitation of LCS that it cannot differentiate LCSs of different spatial relations within their word embedding.
4) **ROUGE-S:** S stands for skip-bigrams co-occurrence statistics. Skip-bigrams are bigrams that do not have to appear together in a sentence. For the sentence I am Ram , the skip-bigrams generated will be (I , am), (I , Ram), (am , Ram) [21].

B. GENERIC PERFORMANCE METRICS

1) **PRECISION:** It is computed by dividing the number of sentences common in the Reference and Candidate summary by the number of sentences in the candidate summary as shown:

$$\text{Precision} = \frac{N(Sr \cap Sc)}{N(Sc)}$$

where

Sr = Reference summary (the ideal summary containing relevant sentences)
Sc = Candidate summary (the summary generated by a system or algorithm)
N(S) = Number of sentences in summary S

2) **RECALL:** It is computed by dividing the number of sentences common in the Reference and Candidate summary by the number of sentences in the reference summary as shown

$$\text{Recall} = \frac{N(Sr \cap Sc)}{N(Sr)}$$

where

Sr = Reference summary (the ideal summary containing relevant sentences)
Sc = Candidate summary (the summary generated by a system or algorithm)
N(S) = Number of sentences in summary S

3) **F-MEASURE:** It is computed by computing the harmonic mean between precision and recall as shown

$$F_2 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

III. Datasets for summarization

Legal document summarization is an emerging field that benefits from a variety of datasets. This section presents key datasets that aid researchers in developing effective summarization models.

A. CanLII

CanLII (Canadian Legal Information Institute) provides comprehensive access to Canadian legal information, including statutes, regulations, and case law. It serves as a valuable dataset for legal research and summarization, enabling analysis of Canadian legal texts [1].

B. Legal Summarization Dataset

The Legal Summarization Dataset, hosted on GitHub by Laura Manor, contains legal documents paired with human-generated summaries. It is specifically designed for training and evaluating summarization models targeting the unique structure of legal texts [2].

C. Zenodo Dataset

Available on Zenodo, this dataset features a collection of legal documents along with their summaries, aimed at advancing research in legal text summarization [3].

D. ILDS

The ILDS repository by SATYAJIT1910 focuses on Indian legal documents and their summaries. This dataset is significant for researchers interested in summarization methods applicable to the Indian legal system [4].

E. AustLII

AustLII (Australian Legal Information Institute) offers free access to Australian legal materials, providing a rich dataset for exploring legal summarization in the context of Australian law [5].

F. Law Reports

Law Reports provide access to various legal reports and case decisions from the UK. Datasets derived from these reports are instrumental for summarizing case law in the UK legal context [6].

G. EUR-Lex Summary Dataset

The EUR-Lex Summary Dataset, curated by Dennlinger, contains summaries of legal texts from the European Union. It is particularly useful for developing summarization techniques specific to EU legal texts [7].

IV. Classification of Summarization Appoaches

As discussed previously, text summarization for legal documents is divided into 3 main types:

1) Extractive approach
2) Abstractive approach
3) Hybrid approach

A. *Extractive Summarization*

Extractive summarization techniques operate by identifying and selecting the most important sentences or phrases directly from the legal document without modifying or generating new text. The primary goal of extractive methods is to retain the core content of the document while minimizing redundancy. This approach is particularly useful in legal contexts where precise language and accuracy are critical. Numerous state-of-the-art methods have been proposed to perform extractive summarization on legal documents, each attempting to handle the inherent challenges of legal language, such as complexity, formality, and domain-specific jargon. Below is a review of several key extractive summarization approaches

1) Extractive Summarization Techniques:

a. Incorporating Domain Knowledge for Extractive Summarization of Legal Case Documents:
DELSumm is an unsupervised summarization algorithm that effectively integrates expert guidelines into the summarization process of legal case documents. Traditional summarization methods often overlook domain-specific knowledge, which is crucial for generating comprehensive summaries. This paper addresses this gap by employing Integer Linear Programming (ILP) to optimize the selection of sentences, ensuring balanced representation of various rhetorical segments such as facts and judgments. Experiments conducted on Indian Supreme Court case documents demonstrate that DELSumm significantly outperforms both general and legal-specific summarization algorithms, achieving superior ROUGE scores. This innovative approach showcases the potential of domain-adaptive summarization methods in enhancing the quality of legal summaries, indicating that incorporating expert knowledge can yield more informative and relevant outputs than traditional supervised models, even with fewer training examples. Overall, this work represents a significant advancement in the field of legal document summarization. [8]

b. Indian Legal Text Summarization: A Text Normalisation-based Approach:
The paper addresses the pressing issue of the Indian judiciary's backlog, with over 4 crore pending cases, highlighting the inefficiency of manual legal document summarization. The authors propose a text normalization methodology to enhance the performance of state-of-the-art models like BART and PEGASUS for summarizing Indian legal texts. This approach aims to streamline the summarization process, making it more efficient for legal professionals and accessible to the public.
The methodology does not rely on publicly available datasets, which are currently lacking. It involves normalizing legal texts to facilitate better summarization through both extractive and abstractive techniques. The authors review existing research, noting that most studies focus on jurisdictions outside India, emphasizing the need for a tailored approach due to unique legal terminology and structures.
Evaluation using ROUGE metrics and expert feedback shows significant improvements in summarization performance, indicating the methodology's potential to alleviate the backlog in the Indian judiciary. The authors conclude that automating summarization can save time for legal professionals and enhance public access to legal information. [9]

c. Summarizing Legal Regulatory Documents using Transformers:
The paper focuses on extractive summarization of legal texts, utilizing a BERT-like architecture to generate sentence representations from source documents. The methodology involves training models to classify sentences as summary candidates through binary classification, where each sentence in a document d is assigned a label y 0,1 indicating whether it should be included in the summary. The authors introduce the EUR-LexSum dataset, which consists of 4,595 English summaries of EU legal acts, structured for a non-specialist audience. This dataset aims to facilitate legal document summarization, addressing the complexity and abundance of legal texts . Additionally, the paper contrasts its approach with prior works that primarily focused on abstractive summarization methods, emphasizing its unique contribution to extractive summarization in the legal domain. [10]

d. LEGAL-BERT: The Muppets straight out of Law School:
The paper introduces LEGAL-BERT, a specialized adaptation of the BERT model for the legal domain, emphasizing the need for tailored strategies due to the unique characteristics of legal text. The authors propose three main techniques for applying BERT:using the original BERT model without modifications,further pre-training BERT on domain-specific legal corpora to enhance its understanding of legal language, and training a new BERT model from scratch on legal texts. This systematic investigation aims to improve performance on legal tasks by leveraging the nuances of legal syntax and semantics. Key findings indicate that further pre-training or training from scratch on legal data significantly outperforms using BERT out of the box. Additionally, a broader hyper-parameter search can enhance performance, and smaller BERT-based models can compete effectively with larger models in specialized domains, promoting efficiency and sustainability [11]

e. Processing Long Legal Documents with Pre-trained Transformers:-Modding LegalBERT and Longformer:
The paper explores two primary techniques to address the challenge of summarizing and classifying long legal documents. The first technique modifies Longformer, which uses sparse attention to handle long texts (up to 4,096 sub-words). This model is further adapted to process even longer documents (up to 8,192 sub-words) by extending positional embeddings and reducing the local attention window. The second technique adapts LegalBERT by incorporating TF-IDF representations, allowing it to pro-

cess longer texts without increasing the model size. While the TF-IDF-based approach improves computational efficiency, the Longformer extension yields better overall performance.

The models were evaluated on LexGLUE's legal NLP benchmark, specifically focusing on long-document classification tasks like SCOTUS and ECtHR. [12]

f. Study of Question Answering on Legal Software Document using BERT based models:

The paper explores the performance of transformer-based models like BERT, ALBERT, LEGAL-BERT, RoBERTa, and DistilBERT in the domain of legal software documents. The focus is on the PolicyQA dataset, a specialized dataset containing privacy policies, and comparing these models' performance with the general-purpose SQuAD V2.0 dataset. Surprisingly, general-purpose models like ALBERT and BERT outperformed domain-specific models like LEGAL-BERT on the PolicyQA dataset. This indicates that legal documents in software development may not benefit as much from domain-specific models due to the complexity and specific language used in privacy policies.

The experiments were conducted over 5 and 10 epochs, with performance measured using Exact Match (EM) and F1 scores. The results showed that the models performed significantly better on the SQuAD V2.0 dataset than PolicyQA, with ALBERT performing the best overall. The authors recommend training models from scratch on legal software texts to improve performance. [13]

g. CaseSummarizer: A System for Automated Summarization of Legal Texts:

This paper presents CaseSummarizer, an automated tool designed for summarizing legal documents. Legal professionals often deal with extensive amounts of text, making it challenging to manage cases effectively. CaseSummarizer employs standard summarization methods based on word frequency, enhanced with domain-specific knowledge, to create concise summaries. The tool features an informative interface that includes elements like abbreviations and significance heat maps, making it user-friendly. The effectiveness of CaseSummarizer is evaluated using the ROUGE metric and human scoring against other summarization systems, demonstrating its capability to provide a reasonable context of legal cases despite some limitations in capturing all important points [14]

h. LetSum, an automatic Legal Text Summarizing system:

The paper discusses the development of LetSum, a system designed for automatic summarization of legal judgments. Its primary goal is to create concise summaries that help legal experts manage large volumes of documents efficiently. The methodology involves several key phases: Thematic Segmentation: The system first breaks down the judgment into thematic sections, such as Introduction, Context, Juridical Analysis, and Conclusion. Filtering: It removes less important content, like citations of law articles, to focus on the main ideas. Selection of Relevant Textual Units: The system identifies and selects the most relevant sentences for each theme. Summary Production: Finally, it compiles these selected sentences into a coherent summary that adheres to size limits 89. The summarization method employed by LetSum is primarily extractive. It extracts key sentences from the original text without altering them, aiming to reproduce the reasoning of human experts while organizing the information effectively 10.

In summary, LetSum combines thematic analysis and filtering techniques to produce extractive summaries that enhance the readability and coherence of legal documents. [15]

2) Evaluation and Results of Extractive Approaches: Extractive summarization has proven effective in several legal document summarization tasks due to its simplicity and ability to preserve factual correctness. However, these methods often lack flexibility in generating more abstract summaries that could convey higher-level legal reasoning. Evaluations typically measure performance using metrics such as ROUGE and BLEU, but these evaluations may not fully capture the importance of legal-specific features like argumentation or case precedents.

For instance, LexRank and graph-based methods show high precision in summarizing court judgments, where the structure of arguments is critical. Reinforcement learning-based extractive methods have recently shown promise by dynamically adjusting summaries based on task-specific reward functions. Additionally, transformer-based models, fine-tuned on legal corpora, represent the current cutting edge in extracting semantically relevant content from complex legal texts.

B. Abstractive Summarization

Abstractive summarization is significantly more complex than extractive summarization because it involves generating new sentences that convey the same meaning as the original text, rather than simply selecting and reordering the existing sentences. In the context of legal documents, this approach requires advanced natural language generation techniques, since legal language is precise and formal, often containing terms and constructs that must be accurately translated into more general or succinct summaries. Here, we review several state-of-the-art abstractive summarization approaches that have been applied to legal documents.

1) Abstractive Techniques:

a. Building an Optimized algorithm that provides summaries of legal documents:

This paper analyzes various NLP algorithms for text summarization, specifically focusing on fine-tuning the T5 model to enhance the accuracy of legal document summaries. Given the extensive documentation in the legal field, the developed algorithm aims to provide accessible and precise summaries for professionals such as lawyers and clients. The paper highlights the importance of these summaries in referencing similar cases and presents a user-friendly interface for inputting documents and generating summaries that can be easily copied by users. This novel approach moves beyond traditional extractive methods by offering paraphrased phrases that utilize a different vocabulary set, making the summaries more relevant and useful [16]

b. Demystifying Legalese: An Automated Approach for Summarizing and Analyzing Overlaps in Privacy Policies and Terms of Service:
This paper focuses on simplifying complex Privacy Policies and Terms of Service (ToS) documents using automated machine learning techniques. The authors developed models, particularly RoBERTa, to extract key concepts and generate summaries that make these documents easier to understand for users. The study also analyzes overlaps between privacy policies and ToS, identifying areas of redundancy or non-compliance with GDPR guidelines. The ultimate goal is to help users make more informed decisions about their data privacy by offering concise summaries and insights into these often confusing legal documents. [18]

2) Evaluation and Results of Abstractive Approaches: Abstractive summarization has made significant strides in recent years due to advances in transformer-based architectures. The legal domain, in particular, has benefited from fine-tuning pre-trained models on legal corpora. Transformer models like BART and T5 excel at generating fluent and coherent summaries but sometimes struggle with legal precision. On the other hand, models like Longformer and Hierarchical Attention Networks specifically target the challenges of handling long legal documents.

Evaluations of abstractive approaches typically use ROUGE, METEOR, and human evaluations, focusing on the accuracy, coherence, and legal soundness of the generated summaries. Despite their advances, abstractive methods still face challenges in maintaining the legal factuality of summaries, and hybrid approaches are often employed to strike a balance between extractive and generative models.

C. Hybrid Summarization Approaches

Hybrid approaches combine both extractive and abstractive summarization techniques to generate more comprehensive and accurate summaries. In legal document summarization, hybrid models often leverage extractive methods to identify the most relevant parts of the document and abstractive techniques to refine and rephrase the extracted text. This ensures the generation of summaries that are both faithful to the original content and concise, avoiding the limitations inherent in using either method in isolation. Below, we discuss several state-of-the-art hybrid summarization methods applied to legal texts.

1) Hybrid techniques:

a. Leveraging dense retrieval and summarization-based reranking for case law retrieval:
This paper focuses on retrieval-based question answering in the legal domain, which is challenging due to the complexity and variety of legal documents. It emphasizes the importance of effective text representation, as better representations lead to more accurate matches between legal questions and articles. The proposed model utilizes neural attentive representations to extract significant information from both questions and legal articles, leveraging convolutional neural networks and attention mechanisms. Experimental results demonstrate that this model significantly outperforms existing state-of-the-art methods, improving recall and NDCG metrics. The study highlights the necessity of retrieving relevant legal articles to facilitate easier access to useful information rather than requiring users to sift through lengthy documents [17]

b. Keyword-based Augmentation Method to Enhance Abstractive Summarization for Legal Documents:
This paper addresses the challenges of summarizing lengthy legal documents, which traditional machine learning models struggle with. The authors propose a keyword-based augmentation approach to improve abstractive summarization. By incorporating important keywords into the summarization process, the model can better focus on key information in legal texts. The study compares different Transformer-based models, such as BART and Longformer Encoder-Decoder (LED), and explores how the quality of extracted keywords impacts the performance of the summarization. The results demonstrate that using high-quality keywords significantly enhances the summarization process, especially for long legal documents. [19]

2) Evaluation and Results of Hybrid Approaches: Hybrid approaches to legal document summarization strike a balance between the extractive and abstractive paradigms. By leveraging extractive techniques to preserve factual accuracy and abstractive techniques to enhance fluency, hybrid models provide comprehensive summaries that are legally accurate and linguistically coherent. The results of these approaches have been consistently positive across various benchmarks, including ROUGE, BLEU, METEOR, and human evaluations.

While hybrid approaches tend to outperform standalone extractive or abstractive methods, they also present challenges in terms of computational complexity and model interpretability. Nonetheless, the adaptability of these methods makes them well-suited for the legal domain, where maintaining precision while generating concise summaries is critical.

V. Analysis of Different Techniques and Discussion

In the context of summarizing legal documents, the choice among extractive, abstractive, and hybrid summarization techniques plays a crucial role in determining the effectiveness of the summaries produced. Each method presents distinct advantages and challenges that merit careful consideration, particularly regarding accuracy, coherence, computational demands, and suitability for legal contexts.

A. Extractive Techniques

Extractive summarization methods are particularly adept at maintaining the integrity and factual correctness of legal

Algorithm	Rouge 2		Rouge L	
	R	F	R	F
DELSumm	0.4323	0.4217	0.6831	0.6017
Indian Legal Text Summarization	0.24	0.25	0.3	0.31
EUR-LexSum(Oracle)	24.912	24.448	25.005	24.314
LEGAL-BERT	0.367	0.347	0.487	0.465
T5		0.175		
GPT-2		0.075		
BERT		0.2		
BART(baseline)	0.555	0.16	0.604	
Pegasus		0.14		
TextRank		0.15		
CaseSum		0.114	0.061	
LED(baseline)	0.71		0.737	
LetSum	0.314		0.452	

TABLE I
SCORES FOR VARIOUS ALGORITHMS ON RAW TEXT.

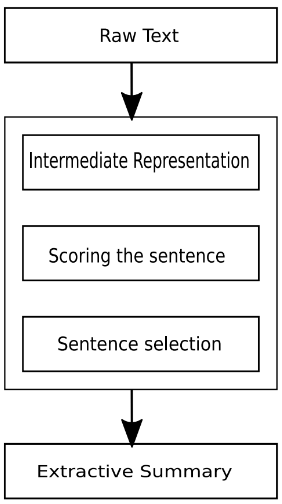

Fig. 1. Extractive text summarization model
[20]

texts. Techniques like Latent Semantic Analysis (LSA) and graph-based approaches such as LexRank are effective at pinpointing key sentences that convey essential arguments. However, these methods can lead to redundancy, often selecting multiple sentences that express similar ideas, resulting in longer summaries that may lack brevity. Additionally, the inability to generate novel phrasing can limit the adaptability of extractive techniques, especially in summarizing complex legal reasoning.

B. Abstractive Techniques

On the other hand, abstractive summarization leverages advanced models like BART and T5 to produce more cohesive and engaging summaries. These approaches can synthesize information and present overarching insights, which is especia lly valuable in legal scenarios requiring nuanced interpretation. Nonetheless, they encounter challenges related to legal accuracy; the generation of new sentences can sometimes lead to misinterpretations of legal concepts. Innovations such as Longformer aim to address issues related to the length of legal

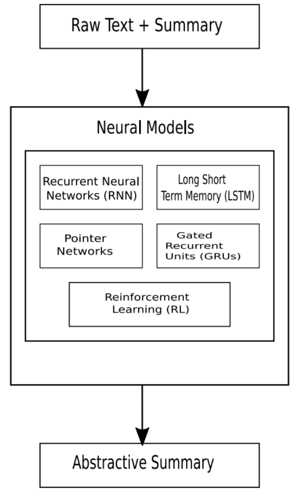

Fig. 2. Abstractive text summarization model
[20]

documents, enabling better management of extensive texts, yet the risk of producing legally inaccurate summaries persists.

C. Hybrid Techniques

Hybrid summarization methods offer a promising compromise by integrating the benefits of both extractive and abstractive techniques. For example, the Extract-then-Abstract framework utilizes an initial extractive phase to identify significant sentences, followed by an abstractive phase to enhance clarity and coherence. This dual-stage process not only improves fluency but also ensures that vital legal terms are retained. Hybrid models have consistently demonstrated superior performance across various evaluation metrics, highlighting their potential to meet the specific requirements of legal summarization. However, they may involve increased complexity in training and greater computational costs.

D. Comparative Performance and Future Directions

While traditional evaluation metrics like ROUGE and BLEU provide valuable insights into summarization quality, they may not fully capture critical legal features, such as the flow of arguments and relevance to precedents. Future research should aim to develop tailored evaluation methodologies that consider these aspects, as well as human-centered assessments to evaluate the practicality and interpretability of the summaries in real-world legal settings.

As advancements in natural language processing continue to unfold, incorporating emerging techniques such as reinforcement learning and multi-task learning may further enhance summarization capabilities. These approaches present opportunities to create models that adapt to the complexities of legal language, thereby improving both the accuracy and fluency of the resulting summaries.

VI. Conclusion

Text summarization has remained a vibrant area of research for the past two decades, with significant advancements made in extractive summarization techniques. However, the automatic generation of abstractive summaries, particularly in the context of legal documents, is still an emerging field that requires further exploration. Legal texts present unique challenges, including their length, specialized terminology, and structural complexity, which differ markedly from other genres.

This survey has provided a comprehensive overview of various summarization techniques, starting with foundational definitions and progressing through state-of-the-art methods. We emphasized the particularities of legal document summarization, addressing critical issues within the domain and categorizing techniques based on approaches such as citation-based, graph-based, and rhetorical roles. Despite the notable achievements in extractive summarization, our review revealed a concerning lack of research into abstractive methods and multi-document summarization within the legal field.

The findings indicate a pressing need for further investigation into novel summarization techniques tailored for legal texts. Additionally, the development of benchmark datasets and evaluation metrics specific to this domain is crucial for enabling comparative analyses and fostering advancements. The exploration of automatic categorization of similar court cases could greatly enhance the efficiency and effectiveness of legal practitioners, providing them with holistic overviews and streamlined access to pertinent information.

In summary, while considerable strides have been made, the journey towards effective legal text summarization is far from complete. Addressing the identified research questions and pursuing the outlined future directions will be essential in advancing this important field.

References

[1] CanLII, "Canadian Legal Information Institute," available: https://www.canlii.org/en/.

[2] L. Manor, "Legal summarization dataset," available: https://www.github.com/lauramanor/legal_summarization.

[3] Zenodo, "Legal summarization dataset," available: https://www.zenodo.org/records/4595826.

[4] SATYAJIT1910, "Indian legal document summarization dataset," available: https://github.com/SATYAJIT1910/ILDS.

[5] AustLII, "Australian Legal Information Institute," available: https://www.austlii.edu.au/.

[6] Law Reports, available: https://www.lawreports.co.uk/.

[7] Dennlinger, "EUR-Lex summary dataset," available: http://huggingface.co/datasets/dennlinger/eur-lex-sum.

[8] Paheli Bhattacharya,Soham Poddar,Koustav Rudra, "Incorporating Domain Knowledge for Extractive Summarization of Legal Case Documents" available: https://arxiv.org/pdf/2106.15876.

[9] Satyajit Ghosh,Mousumi Dutta,Tanaya Das "Indian Legal Text Summarization: A Text Normalisation-based Approach," available:https://arxiv.org/pdf/2206.06238.

[10] Svea Klaus,Ria Van Hecke,Kaweh Djafari Naini,Ismail Sengor Altingovde,Juan Bernab-Moreno,Enrique Herrera-Viedma ,"Summarizing Legal Regulatory Documents using Transformers" available: https://web.archive.org/web/20220712231042id_/https://dl.acm.org/doi/pdf/10.1145/3477495.3531872

[11] Ilias Chalkidis , Manos Fergadiotis ,Prodromos Malakasiotis , Nikolaos Aletras,Ion Androutsopoulos, "LEGAL-BERT: The Muppets straight out of Law School," available:https://aclanthology.org/2020.findings-emnlp.261.pdf

[12] Dimitris Mamakas,Petros Tsotsi,Ion Androutsopoulos,Ilias Chalkidis "Processing Long Legal Documents with Pre-trained Transformers:Modding LegalBERT and Longformer," available:https://arxiv.org/pdf/2211.00974v2

[13] Ernesto Quevedo Caballero, Mushfika Rahman, Tomas Cerny,Pablo Rivas, and Gissella Bejarano, "Study of Question Answering on Legal Software Document using BERT based models available:https://research.latinxinai.org/papers/naacl/2022/pdf/paper_09.pdf

[14] Seth Polsley Pooja Jhunjhunwala,Ruihong Huang, 'CaseSummarizer: A System for Automated Summarization of Legal Texts," available:https://people.engr.tamu.edu/huangrh/papers/coling16_caseSummarizer.pdf

[15] Atefeh Farzindar,Guy Lapalme, "Letsum, an automatic legal text summarizing system," available:https://www.researchgate.net/publication/228980166_Letsum_an_automatic_legal_text_summarizing_system

[16] Aman Burman1,Eric Bradford, "Building an Optimized algorithm that provides summaries of legal documents," available:https://www.researchgate.net/publication/365294672_Building_an_Optimized_algorithm_that_provides_summaries_of_legal_documents

[17] SHIKHA SONEJI, MITCHELL HOESING, SUJAY KOUJALGI, and JONATHAN DODGE, "Demystifying Legalese: An Automated Approach for Summarizing and Analyzing Overlaps in Privacy Policies and Terms of Service," available:https://arxiv.org/pdf/2404.13087

[18] Sophia Althammer,Arian Askari,Suzan Verberne,Allan Hanbury, "Leveraging dense retrieval and summarization-based re-ranking for case law retrieval," available:https://arxiv.org/pdf/2108.03937

[19] Huyen Nguyen,Junhua Ding,"Keyword-based Augmentation Method to Enhance Abstractive Summarization for Legal Documents," available:https://dl.acm.org/doi/pdf/10.1145/3594536.3595120

[20] Deepali Jain, Malaya Dutta Borah, Anupam Biswas, Summarization of legal documents: Where are we now and the way forward, Computer Science Review, Volume 40, 2021, 100388, ISSN 1574-0137, https://doi.org/10.1016/j.cosrev.2021.100388.

[21] D. Yadav, R. Katna, A. K. Yadav and J. Morato, "Feature Based Automatic Text Summarization Methods: A Comprehensive State-of-the-Art Survey," in IEEE Access, vol. 10, pp. 133981-134003, 2022, doi: 10.1109/ACCESS.2022.3231016