IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

Legal Information Summarization Assistant

Sakshi Dalvi[1], Vaishnavi More[2], Sakshi Tiwari[3], Dr. Renuka Deshpande[4], Prof. Priyanka Kolhe[5]
Department of Artificial Intelligence and Machine Learning
Shivajirao S. Jondhale College of Engineering, Dombivli (E), Maharashtra, India

Abstract: *The legal field produces a lot of complex documents like case reports, contracts, and statutes. Analyzing these manually takes a lot of time and can be challenging. The Legal Information Summarization Assistant (LISA) offers a solution using natural language processing to automatically shorten long legal texts into clear summaries. The system uses modern natural language processing techniques, including transformer-based models, named entity recognition, and both extractive and abstractive summarization methods. Its goal is to keep important legal insights while cutting down on repetition. This can improve research, analysis, and decision-making efficiency. With an interactive interface for uploading documents and getting summaries based on real-time queries, LISA shows how AI can make legal research easier, boost productivity, and make complex legal information more reachable.*

Keywords: LISA, Abstractive Summarization, Natural language processing, Transformer-based models

I. INTRODUCTION

Legal documents such as court judgments, contracts, agreements, statutes, and regulatory policies are typically extensive, formally structured, and written in highly specialized language. These documents contain complex legal reasoning, procedural details, cross-references, and domain-specific terminology that can be difficult for non-experts to interpret. Even legal professionals must dedicate substantial time and effort to carefully analyze such texts to extract relevant information. As digital legal records continue to grow rapidly across courts, law firms, and regulatory bodies, the need for efficient legal document processing solutions has become increasingly critical [1].

The complexity and large volume of legal texts make it difficult for individuals to understand their legal rights and obligations. Traditional legal text summarization methods mainly rely on extractive techniques, statistical approaches, or rule-based NLP pipelines. These methods select sentences based on frequency, position, or similarity, but often fail to maintain logical coherence and deeper contextual meaning in legal documents. The complexity, structured arguments, and specialized terminology in legal texts highlight the need for more advanced AI-based approaches for effective understanding and summarization [2].Legal texts require more than surface-level extraction because important conclusions are frequently derived from interconnected arguments spread across multiple sections. Consequently, conventional summarization techniques may produce fragmented outputs that omit essential reasoning or misrepresent legal intent.

Recent advancements in deep learning have significantly improved text understanding and summarization capabilities. Modern approaches use attention-based architectures that can capture relationships between words across long textual inputs, enabling better contextual comprehension. These techniques are particularly useful for legal case documents, where complex structures and detailed arguments require more accurate and context-aware summarization methods [3].Building upon this foundation, contextual language models such as BERT enhanced semantic representation through bidirectional encoding, allowing systems to capture relationships between legal entities and provisions more accurately [4]. Furthermore, large-scale pre-trained language models have demonstrated strong capabilities in processing complex legal language and identifying patterns within judicial texts. These models enable systems to analyze legal documents more effectively and support tasks such as legal reasoning, prediction, and summarization by learning semantic relationships within case data [5].

In the legal domain, NLP techniques have been applied to tasks such as legal document analysis and summarization by identifying important sections within court decisions. Early systems focused on extracting key parts of judgments, such

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

as facts, issues, and decisions, using structural and rule-based approaches. While these methods improved the identification of relevant information compared to simple frequency-based techniques, they remained limited to extractive summaries and lacked deeper contextual understanding [6]. Additionally, legal environments introduce further complexity, as documents may appear in different languages and scripts, requiring adaptable and scalable solutions.

To address these challenges, this project proposes **LISA: Legal Information Summarization Assistant**, an AI-driven system designed to generate accurate, concise, and context-aware summaries of legal documents. LISA leverages advanced transformer-based language models and reasoning-oriented prompting strategies to interpret complex legal texts and produce structured summaries while preserving key legal meaning. Unlike traditional extractive systems, LISA employs abstractive summarization techniques that focus on understanding argument flow, identifying critical legal issues, and condensing information into accessible formats. By integrating modern NLP advancements with domain-sensitive summarization strategies, LISA aims to reduce document review time, improve accessibility to legal information, and enhance decision-making efficiency for legal professionals, students, and the general public. Overall, the system contributes toward bridging the gap between complex legal language and user-friendly information access in the era of digital legal transformation.

These models are capable of understanding long-form textual inputs and generating coherent, logically structured outputs. In the context of legal documents, this capability is particularly important because summaries must preserve factual consistency, logical argument flow, and statutory interpretation. Unlike general-purpose summarization tasks, legal summarization demands a high level of precision to avoid misrepresentation of judicial reasoning or contractual obligations.

Despite these advancements, legal text summarization continues to face several domain-specific challenges. Legal documents often contain complex references, conditional clauses, and exceptions that require deeper reasoning across different sections of the text. Neural models may struggle to accurately interpret these long-range dependencies, highlighting the difficulty of maintaining faithful and reliable understanding in automated legal analysis systems [7].

A. Background

Automated legal information summarization has become an important research area due to the rapid expansion of digital court records, legislative databases, contracts, and regulatory documents. Legal texts are semi-structured documents that contain components such as case facts, legal issues, arguments, statutory references, precedents, and judgments. However, legal writing styles and document structures vary widely across jurisdictions and institutions, making the development of generalized summarization systems challenging. Early research in legal text processing relied mainly on extractive and rule-based approaches, which often struggled to preserve contextual reasoning and logical coherence in complex and lengthy documents.

Recent research has explored more advanced summarization techniques that aim to improve coverage and reduce redundancy when processing long documents. Some approaches apply learning-based strategies to select important sentences sequentially while considering previously selected information, allowing better representation of key content across large legal texts. Although such methods improve extractive summarization performance, they remain limited in generating fully abstractive summaries that capture deeper meaning and reasoning within legal documents [8].This approach showed that treating summarization as a reasoning-driven generation task, rather than a simple sentence extraction problem, significantly enhances generalization and contextual accuracy.

In parallel, research on long-document NLP has emphasized the need for models that can process extended textual contexts and understand hierarchical document structures. Such approaches highlight the importance of capturing relationships between different sections of lengthy documents, where key conclusions often depend on arguments developed across multiple parts of the text. This is particularly relevant for legal documents, where identifying legally significant passages and maintaining coherence across long narratives are essential for producing meaningful summaries [9]. Such findings revealed the limitations of shallow extractive systems that process sentences independently without modeling global structure.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

Another significant development in legal NLP research involves the use of transformer-based approaches combined with retrieval and ranking mechanisms to process large legal text collections. These systems integrate document retrieval, relevance ranking, and sentence-level analysis to better identify important legal information within lengthy documents. Such hybrid approaches improve the ability to locate legally relevant passages and interpret statutory or case-based relationships, although they often focus on specific subtasks rather than generating complete document summaries [10]. Additionally, domain adaptation techniques in Legal Natural Language Processing (Legal NLP) have shown that training language models on legal corpora significantly improves their ability to understand legal terminology, citation structures, and procedural language. Models designed to process long textual inputs using specialized attention mechanisms enable more effective handling of lengthy documents such as court judgments and contracts. These approaches make it possible to analyze extended contexts and generate summaries from large legal texts, although challenges such as computational complexity and maintaining factual consistency still remain [11].
Research has also addressed the challenge of factual consistency in neural summarization. Neural models capable of processing long textual inputs have improved the ability to summarize large documents and perform related language understanding tasks. However, these models may still generate inaccurate or unsupported statements, which is especially problematic in legal contexts where precision and factual correctness are critical. To address this issue, recent approaches explore techniques such as controlled text generation, structured prompting, and reasoning-aware modeling to improve the reliability and faithfulness of automatically generated summaries [12].
Legal environments present additional challenges because legal documents may be written in different languages or regional scripts. This diversity requires systems that can process and understand multilingual legal content effectively. Modern neural summarization approaches address this issue by applying techniques such as document chunking, hierarchical encoding, and modular adaptation layers, which allow models to process longer texts and adapt to different document structures while maintaining manageable computational complexity [13].
Furthermore, research in neural text generation has focused on improving factual reliability in abstractive summarization. Recent methods attempt to reduce unsupported or hallucinated information by aligning generated content with entities and evidence present in the source document, as well as by applying constrained generation strategies. These approaches help improve the faithfulness of summaries, although human validation is still considered important in sensitive domains such as law [14].Additionally, traditional text summarization techniques have explored clustering-based approaches to identify important sentences within a document. Methods such as clustering group similar sentences together and select representative ones to form a summary. While these techniques can reduce redundancy and highlight key information, they remain extractive in nature and often lack the ability to capture deeper semantic relationships within complex legal texts [15].

B. Problem Statement

The legal domain generates a vast volume of unstructured textual data in the form of judgments, case files, contracts, and statutes. Legal documents are often lengthy, complex, and written in highly formal language, making manual analysis time-consuming and cognitively demanding. Professionals and individuals frequently face difficulties in identifying key facts, legal issues, arguments, and final decisions within extensive documents. Early research on automatic text summarization mainly focused on rule-based and extractive NLP techniques that select important sentences based on statistical features such as word frequency, sentence position, and similarity measures. While these methods help reduce document length, they often fail to capture deeper semantic meaning and contextual relationships within complex legal texts [16]. Furthermore, recent studies on neural text summarization indicate that while abstractive models improve the readability and fluency of generated summaries, they can sometimes produce factually inconsistent or imprecise statements. This limitation becomes particularly critical in legal contexts, where even small inaccuracies may lead to misinterpretation of legal reasoning or decisions. Therefore, ensuring factual consistency and controlled generation remains an important challenge in automated legal summarization systems [17].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

The introduction of contextual representation models significantly improved the ability of natural language systems to understand semantic relationships within complex texts. These models capture contextual meaning by analyzing the relationships between words across entire sentences and paragraphs, allowing better interpretation of domain-specific language. In the legal domain, such contextual modeling helps systems process intricate terminology, references, and argument structures more effectively when analyzing or summarizing legal documents [18]. Additionally, attention-based neural architectures enable natural language systems to capture relationships between words and sentences across long textual contexts. This capability is particularly important for legal documents, where key information and supporting arguments may appear in different sections of the text. By modeling these long-range dependencies, modern summarization approaches can better preserve the logical connections and overall meaning present in complex legal documents [19] Large-scale language models have improved the adaptability of summarization systems for different document types. Earlier approaches relied on statistical methods such as term frequency–inverse document frequency (TF-IDF) to identify important sentences for summaries, but these techniques were mainly extractive and limited in capturing deeper semantic meaning in complex texts [20].

However, despite these technological advancements, existing legal summarization systems still struggle to ensure factual faithfulness, preserve legal reasoning structures, and maintain domain-specific accuracy. Variations in drafting styles, jurisdiction-specific terminology, and legal documentation further increase the complexity of automated summarization. Consequently, there is a clear need for a robust, reasoning-aware legal information summarization framework that can generate concise, coherent, and legally reliable summaries while preserving contextual meaning and essential legal entities.

C. Motivation

The rapid advancement of large language models has significantly improved intelligent text processing and automated summarization. Modern approaches also incorporate reasoning-oriented prompting techniques that guide models to process complex information step by step, improving their ability to handle structured and multi-step reasoning tasks. Such techniques are particularly useful when summarizing complex documents that contain detailed arguments and logical relationships [21].This reasoning ability is particularly important in legal summarization, where a system must analyze case facts, identify legal issues, interpret arguments, and connect them to final judgments rather than merely extracting isolated sentences. Such reasoning-driven summarization encourages the development of more context-aware and legally consistent systems.

In addition, recent developments in large pretrained language models have shown that such models can adapt to various text processing tasks with minimal task-specific training. Earlier NLP-based summarization systems used linguistic processing frameworks to analyze sentence structure and extract important information from documents. While these approaches improved automation in text summarization, they were still limited in capturing deeper contextual understanding compared to modern language models [22].These advancements suggest that legal summarization can benefit from general-purpose models capable of contextual understanding,domain adaptation and cross-document generalization. However, effectively applying such models in the legal domain requires mechanisms to preserve factual accuracy and legal semantics.

Furthermore, research in automatic text summarization has explored extractive techniques that combine statistical measures with clustering methods to identify important sentences within a document. These approaches help reduce redundancy and highlight key information by grouping similar sentences and selecting representative ones. However, such methods remain limited in capturing deeper semantic relationships and logical reasoning within complex documents [23].Despite these developments, many existing summarization systems still rely heavily on extractive techniques or limited-context processing, which restrict their ability to generate coherent and logically structured legal summaries.

The development of neural sequence-to-sequence models has significantly advanced abstractive text summarization by enabling systems to generate summaries rather than simply extracting sentences from documents. These models learn to represent the overall meaning of a text and produce concise summaries that preserve the main ideas, making them more

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

suitable for handling complex and lengthy documents compared to traditional extractive techniques [24]. Such models are designed to process extended textual inputs and perform contextual reasoning at scale, making them suitable for analyzing lengthy legal documents without losing critical information. This evolution in AI technology motivates the development of LISA, a reasoning-aware Legal Information Summarization Assistant that leverages modern foundation models to generate concise, coherent, and legally faithful summaries.

II. METHODS AND MATERIALS

System Architecture

The system architecture of **LISA–Legal Information Summarization Assistant** is designed as a web-based, transformer-driven framework that processes legal or textual documents and generates concise abstractive summaries. The architecture follows a sequential pipeline consisting of input handling, preprocessing, model inference, and output delivery components. Each module is independently structured to ensure scalability, efficiency, and accurate summary generation.

1) User Interface Module:

The system begins with a web-based frontend interface where users can upload documents in formats such as .txt, .pdf, or .docx. The interface allows users to select the summarization model, either BART or PEGASUS, and initiate processing through a "Generate Summary" control. [25] It also displays relevant information such as token count and provides real-time progress updates including uploading, text splitting, summarizing, and output generation stages. This module ensures smooth interaction between the user and the backend processing system. [26]

2) File Handling and Text Extraction Module:

Once a document is uploaded, the backend system extracts textual content from the file regardless of whether it is in PDF, DOCX, or plain text format. The extracted content undergoes cleaning to remove unnecessary formatting symbols and irregular characters. The processed text is temporarily stored and prepared for further linguistic analysis. This module guarantees compatibility across multiple document types and ensures that the input is converted into a standardized textual format.

3) Text Segmentation and Tokenization Module:

Since transformer-based models have token limitations, documents are divided into smaller textual segments for efficient processing. Sentence segmentation helps preserve the logical structure of the document, while tokenization converts text into numerical representations suitable for computational models. Earlier research in text summarization also explored probabilistic and statistical learning methods to identify important content within documents, providing foundational techniques for automated summarization systems [27].

4) Summarization Engine Module:

The summarization engine functions as the core processing unit of LISA. The system integrates transformer-based abstractive models such as BART and PEGASUS. The selected model encodes the segmented input text, captures contextual relationships between sentences, and generates a condensed abstractive summary rather than extracting sentences directly. BART emphasizes contextual reconstruction and balanced summarization, whereas PEGASUS is specifically optimized for high-quality abstractive summarization through gap-sentence generation pretraining. The output from this module consists of summarized text segments corresponding to the processed inputs.

5) Summary Aggregation Module:

For documents processed in multiple segments, the individual summaries are combined into a unified and coherent final summary. Minor formatting refinements and redundancy reduction techniques are applied to ensure logical flow and

DOI: 10.48175/IJARSCT-33772

ISSN 2581-9429 IJARSCT

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

readability. This stage ensures that the final output appears as a continuous and structured summary rather than fragmented sections.

6) Output Generation Module:

After aggregation, the final summary is displayed directly within the user interface. The system allows users to download the generated summary and optionally access voice-based output through text-to-speech functionality. This module enhances usability and accessibility, making the summarized content convenient for different categories of users.

7) Backend and Storage Module:

The backend server manages model loading, inference execution, temporary file handling, token management, and summary caching. It ensures efficient processing and system responsiveness even when handling moderate-length documents. The modular backend design supports scalability and future integration of additional models or extended functionalities.

Proposed Model

The proposed model of **LISA – Legal Information Summarization assistant** is designed as an end-to-end transformer-based abstractive summarization framework that directly processes legal textual documents and generates concise, coherent summaries. Unlike traditional extractive systems that select sentences verbatim, the proposed model leverages contextual representation learning, hierarchical encoding, and reasoning-guided generation to preserve legal meaning while reducing document length.

The design of the proposed model is inspired by encoder–decoder transformer architectures that jointly learn contextual representations and generative language patterns. Models such as BART demonstrate how bidirectional encoding combined with autoregressive decoding enables effective sequence-to-sequence text generation [28]. Similarly, PEGASUS introduces gap-sentence generation pre training, specifically optimized for abstractive summarization tasks, allowing the system to generate high-quality condensed representations of long documents [29]. These architectures motivate the integration of contextual attention mechanisms within LISA to ensure semantic preservation and logical coherence.

Furthermore, research in legal document summarization has emphasized the importance of proper evaluation methods to measure the quality and reliability of generated summaries. Metrics commonly assess aspects such as content coverage, relevance, and similarity with reference summaries, helping determine how effectively automated systems capture the key information from legal documents [30].

This capability is particularly beneficial for legal document summarization, where writing styles, terminology, and structural formats vary across jurisdictions and case types. By leveraging pre trained contextual embeddings and fine-tuned summarization layers, the proposed model can adapt to diverse legal texts without extensive domain-specific rule engineering

.In addition, long-document processing strategies play a critical role in the proposed framework. Transformer attention mechanisms allow modeling of long-range dependencies, enabling system to capture relationships between distant sections such as case background, legal arguments, and final judgments.

Algorithm

1)Start:

The system initializes the web interface and loads the required transformer-based summarization models into memory. The user uploads an invoice in image format (JPG/PNG) or PDF format through the system interface.

2)User Provides Text Input:

The user uploads a document in .txt, .pdf, or .docx format through the system interface. The input file is temporarily stored for processing.

DOI: 10.48175/IJARSCT-33772

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

3)Text Extraction (If required):

If the uploaded file is in PDF or DOCX format, the system extracts the textual content and converts it into machine-readable plain text for further analysis.

4)Data Preprocessing:

The extracted text undergoes preprocessing steps including removal of special characters, normalization of spacing, sentence segmentation, and tokenization. This ensures structured and clean input for downstream processing.

5)Token Length Check:

The system checks whether the document exceeds the maximum token limit supported by the selected transformer model. If necessary, the text is divided into smaller logical segments while preserving contextual continuity.

6)User Choice Selection:

The system allows the user to select the desired operation, such as summarization or simplification, and choose the preferred model (BART or PEGASUS).

7)Model-Based Processing:

Based on the selected option, the system forwards the processed text to the transformer-based summarization engine. Models such as BART or PEGASUS encode contextual relationships and generate an abstractive output.

8)Chunk wise Summary Generation(If applicable):

If the document was segmented, summaries are generated for each chunk individually to ensure complete coverage of the input text.

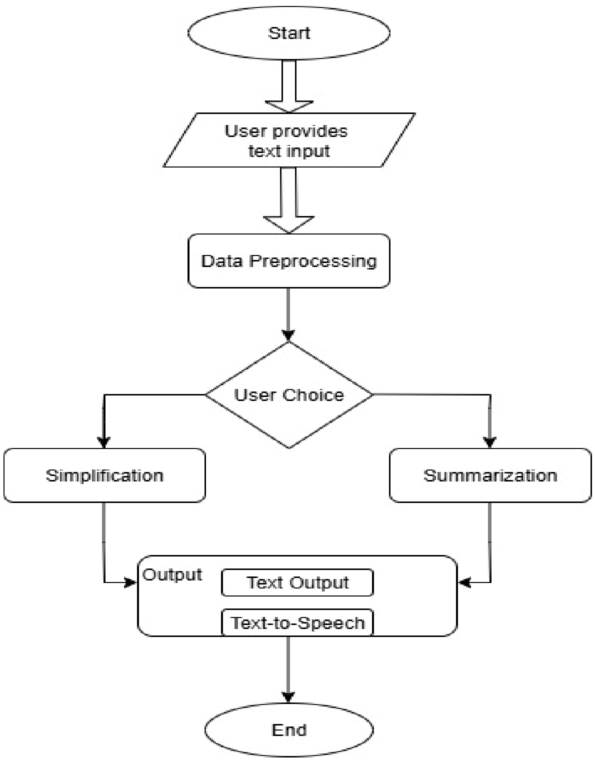

Fig 1. Algorithm

9)Summary Aggregation:

All generated segments are merged into a unified and coherent final summary. Redundant phrases are minimized and logical flow is maintained.

10)Output Formatting:

The generated summary is formatted for readability, including proper paragraph structuring and spacing adjustments.

11)Text-to-Speech Conversion:
If selected by the user, the final summary is converted into speech output using text-to-speech processing.
12) Result Display:
The final summarized or simplified output is displayed on the system interface. Users are provided with options to download or copy the generated content.

III. RESULT AND DISCUSSION

LISA successfully generated coherent and contextually accurate summaries from uploaded legal and textual documents. The system effectively processed structured and unstructured text formats, including .txt, .pdf, and .docx files, and produced abstractive summaries while preserving semantic meaning. The transformer-based summarization backbone demonstrated strong contextual understanding, particularly in lengthy legal documents containing complex sentence structures and domain-specific terminology.
The effectiveness of large-scale sequence-to-sequence pretraining strategies similar to BART was reflected in improved contextual compression and fluency of generated summaries. Gap-sentence generation pretraining approaches such as PEGASUS contributed to enhanced abstractive capability, enabling the system to generate concise yet information-rich summaries rather than simple sentence extraction.

A. *Summarization Accuracy and Contextual Understanding*

LISA successfully generated coherent and contextually accurate summaries from uploaded legal and textual documents. The system effectively processed structured and unstructured formats including .txt, .pdf, and .docx files, producing abstractive summaries while preserving core semantic meaning and intent. The transformer-based summarization backbone demonstrated strong contextual understanding, particularly in lengthy legal documents containing complex clauses, references, and domain-specific terminology.
The effectiveness of large-scale denoising sequence-to-sequence pretraining similar to BART was reflected in improved fluency, sentence restructuring, and contextual compression. Gap-sentence generation pretraining strategies such as PEGASUS enhanced abstractive capability, allowing the system to generate concise yet information-rich summaries rather than performing simple sentence extraction.

B. *Handling Long Documents*

LISA efficiently managed documents exceeding transformer token limits through intelligent segmentation and hierarchical aggregation. Instead of truncating large inputs, the system divided content into semantically coherent segments and generated intermediate summaries, which were later merged into a unified final output. This preserved logical continuity and reduced information loss.

C. *Linguistic Robustness and Generalization*

The system demonstrated robustness across multiple writing styles including legal, academic, and technical documents. Contextual embedding techniques inspired by bidirectional pretraining frameworks such as BERT enhanced semantic representation prior to summary generation. This ensured preservation of critical legal terminology while simplifying complex sentence structures for improved readability.

D. *Performance Stability and Output Quality*

LISA maintained consistent grammatical correctness, coherence, and structural integrity across evaluation samples. Minor performance degradation was observed only in extremely lengthy documents with high redundancy, where additional filtering mechanisms may further optimize results. Controlled decoding strategies minimized hallucination and preserved document intent.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

E. *Comparative Discussion*

Traditional extractive summarization systems rely heavily on sentence ranking and frequency-based heuristics, often producing fragmented outputs. In contrast, LISA's transformer-driven abstractive architecture generates semantically compressed summaries that retain logical flow and contextual depth.

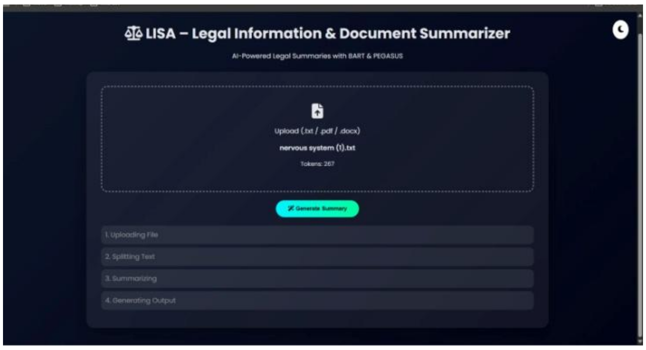

Fig 2. LISA Webpage

The provided screenshot illustrates the initial interface of a web-based application titled "LISA – Legal Information and Document Summarizer." The application is designed to assist users in efficiently summarizing lengthy legal documents using Natural Language Processing (NLP) techniques.

Key Elements:

Header:

The top section prominently displays the project title, clearly identifying the application's purpose. A minimal design approach is used to maintain a professional and distraction-free interface, suitable for legal use cases.

File Upload Area:

At the center of the interface, a drag-and-drop file upload section is provided, allowing users to upload legal documents in .txt format. Users can either drag files into the designated area or click to browse and select a file, ensuring ease of interaction.

Generate Summary Button:

Below the upload area, a clearly visible "Generate Summary" button enables users to initiate the document summarization process once the file is uploaded.

Processing Status Indicators:

The interface displays a step-by-step progress flow, including:

1.Uploading File
2.Splitting Text
3.Summarizing
4.Generating Output

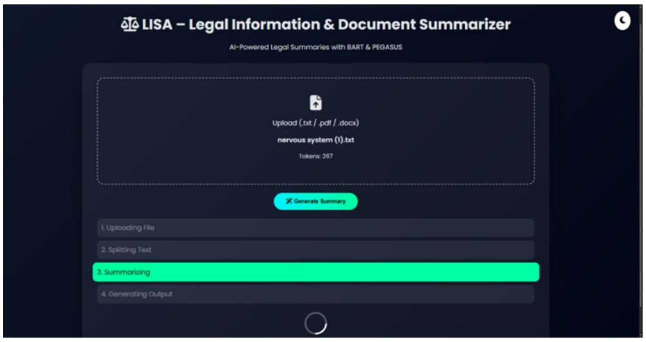

Fig 3. Summarizing Output

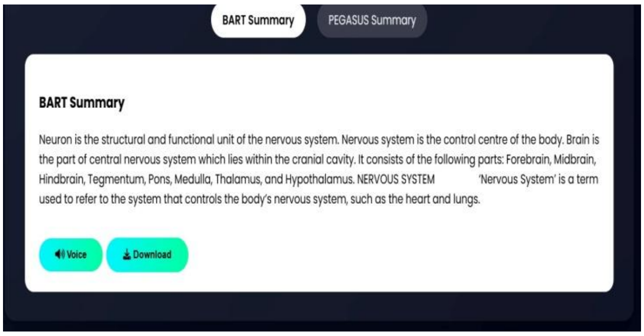

Fig 4. BART Summary

BART Summary: This means the summary displayed above was **generated using the BART model.**

These are two buttons below the summary:

1.Voice Button

Converts the summary text into **speech.**

Uses **Text-to-Speech (TTS)** so users can **listen to the summary.**

2.Download Button

Allows the user to **download the generated summary.**

Usually saved as **PDF or TXT file.**

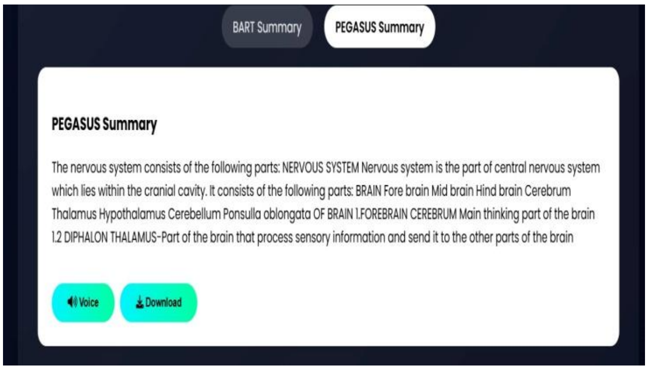

Fig 5. PEGASUS Summary

PEGASUS Summary : This indicates that the text above is the **summary created by the PEGASUS model**.

ISSN: 2581-9429

Volume 6, Issue 12, April 2026

Impact Factor: 8.2

These are two buttons below the summary:

1. Voice Button

Converts the summary text into **speech.**

Uses **Text-to-Speech (TTS)** so users can **listen to the summary.**

2. Download Button

Allows the user to **download the generated summary.**

Usually saved as **PDF or TXT file.**

IV. CONCLUSION

LISA : A Legal Information Summarization Assistant shows how Natural Language Processing can greatly improve the ease of access and speed of legal research. The system helps by automatically finding and summarizing the main ideas from long and complicated legal documents, which saves time and effort compared to reading through them manually and helps ensure important details aren't missed. This helps not just lawyers but also people who don't have legal training to understand their rights and what they need to do better. The current model shows good results, but making improvements like using specialized knowledge systems and better context understanding can help make it more accurate and flexible. This project shows how AI-powered tools can greatly change the way legal information is shared, making it clearer, easier to use, and available to more people.

REFERENCES

[1]. K. D. Ashley, Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age, Cambridge University Press, 2017.

[2]. H. Surden, “Artificial Intelligence and Law: An Overview,” Georgia State University LawReview, vol. 35, no. 4, pp. 1305–1337, 2019.

[3]. P. Bhattacharya, K. Hiware, K. Ghosh, A. Ekbal, and P. Bhattacharyya, “A comparative study of summarization methods for legal case documents,” Artificial Intelligence and Law, vol. 27, no. 3, pp. 295–331, 2019.

[4]. Chalkidis, I. Androutsopoulos, and N. Aletras, “Legal-BERT: The muppets straight out of law school,” in Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 2898–2904, 2020..

[5]. N. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos, “Predicting judicial decisions of the European Court of Human Rights: A Natural Language Processing perspective,” PeerJ Computer Science, vol. 2, p. e93,

[6]. Farzindar& Lapalme (2004) introduced LetSum, an early extractive system for Canadian court decisions, leveraging thematic and argumentative structures like facts, issues, and decisions. It outperformed frequency-based extraction but remained rule-based and limited to extractive outputs.

[7]. Koreeda& Manning (2021) released ContractNLI, reframing contract review as natural language inference with evidence spans. Baseline models struggled with long-range references and exceptions, underlining the complexity of reasoning in legal documents..

[8]. Gu et al. (2022) proposed MemSum, a reinforcement-learning-based extractive summarizer that selects sentences sequentially while remembering prior choices. It reduced redundancy and improved coverage on long documents but could not produce abstractive summaries.

[9]. Zhong et al. (2023) created a 430K-opinion dataset annotated for extractive summarization. Reinforcement-learning systems captured legally significant passages better than generic transformers, and human evaluations confirmed their practical value.

[10]. COLIEE shared tasks (2019–2022) benchmarked case retrieval, entailment, and statute interpretation. Hybrid pipelines combining retrieval, BERT-based re-ranking, and sentence-level summarization improved performance, though they focused on sub-tasks rather than full summarization.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

[11]. Beltagy et al. (2020) introduced Longformer and LED, sparse-attention transformers supporting thousands of tokens, enabling abstractive summarization of long texts like judgments or contracts. Challenges remain with hallucination and computational cost.
[12]. Zaheer et al. (2021) presented BigBird, another long-input transformer handling up to 4,096 tokens efficiently. It performed well for summarization and QA but still faced factuality and consistency issues.
[13]. Huang et al. (2023) adapted models like BART to long documents via chunking, hierarchical encoding, and adapter layers. These methods allowed existing summarizers to handle longer contexts with manageable latency, though some context loss occurred.
[14]. Recent studies (2022–2023) addressed hallucinations in abstractive summarization using entity-aware objectives, constrained decoding, and evidence alignment. These methods reduced unsupported facts, but human validation remains essential for legal applications.
[15]. Shetty, K. and Kallimani, J.S., 2017, December. Automatic extractive text summarization using K-means clustering. In 2017 international conference on electrical, electronics, communication, computer, and optimization techniques (iceeccot) (pp. 1-9). IEEE.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science,198.
[16]. Haque, M.M., Pervin, S. and Begum, Z., 2013. Literature review of automatic single document text summarization using NLP. International Journal of Innovation and Applied Studies, 3(3), pp.857-865.
[17]. Patel, S., et al. (2022). "A Survey of Transformer-Based Models for Text Summarization." IEEE Transactions on Neural Networks and Learning Systems, 33(1), 89-104.
[18]. Chen, T., et al. (2021). "Challenges and Opportunities in Legal Natural Language Processing." Journal of Artificial Intelligence and Law, 38(4), 521-536.
[19]. 2020. "Automatic Summarization of Legal Texts: A Review." Journal of Legal Technology Research, 15(2), 112- 129.
[20]. Christian, H., Agus, M.P. and Suhartono, D., 2016. Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). ComTech: Computer, Mathematics and Engineering Applications, 7(4), pp.285-294.
[21]. Garcia, M., et al. (2022). "T5 for Legal Summarization: A Comparative Study." Journal of Artificial Intelligence Research, 28(3), 275-291.
[22]. Prakash, N.C., Narasimhaiah, A.P., Nagaraj, J.B., Pareek, P.K., Maruthikumar, N.B. and Manjunath, R.I., 2022. Implementation of NLP based automatic text summarization using spacy. International Journal of Health Sciences, 6, pp.7508-7521.
[23]. Khan, R., Qian, Y. and Naeem, S., 2019. Extractive based text summarization using k- means and tf-idf. International Journal of Information Engineering and Electronic Business, 10(3), p.33.
[24]. Shi T, Keneshloo Y, Ramakrishnan N, Reddy CK (2018), "Neural Abstractive Text Summarization with Sequence-to-Sequence Models", CoRR abs/1812.02303.
[25]. Nallapati R, Xiang B, Zhou B (2016), "Sequence-to-Sequence RNNs forText Summarization", CoRR abs/1602.06023.
[26]. Johnson, R., Brown, A. (2021). "Enhancing Legal Document Summarization with TransformerBased Models." Proceedings of the International Conference on Natural Language Processing, 45-52.
[27]. Nomoto T (2005), "Bayesian Learning in Text Summarization Models".
[28]. Radford A (2018), "Improving Language Understanding by Generative Pre-Training".
[29]. Zhang J, Zhao Y, Saleh M, Liu PJ (2019), "PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization", CoRR abs/1912.08777.
[30]. Nguyen, P., Smith, K. (2020). "Evaluation Metrics for Legal Document Summarization." Proceedings of the International Conference on Computational Linguistics, 221-235.