

Jurisynt : AI for Legal Clarity and Efficiency

Prof. Jordan Choudhari¹, Prathamesh Holay², Gargi Thipase³, Pranita Panchal⁴, Shrikant Jadhav⁵

Prof. Jordan Choudhari,
Prathamesh Holay,
Gargi Thipase,
Pranita Panchal,
Shrikant Jadhav,

Computer Science and Engineering,
Computer Science and Engineering,
Computer Science and Engineering,
Computer Science and Engineering,
Computer Science and Engineering,

Nutan College Of Engineering and Research
Nutan College Of Engineering and Research
Nutan College Of Engineering and Research
Nutan College Of Engineering and Research
Nutan College Of Engineering and Research

Abstract - The number of legal documents is growing so fast and this has created a very big need for ways to analyze the legal text that are automatic which can handle a lot of work and get better understanding of the text. The growth of legal databases has created this demand that motivated us to represent this paper JURISYNTH, an intelligence framework that used advanced Natural Language Processing techniques to make the legal documents simpler and easier to summarize and predict legal outcomes, contractual risks with high accuracy and case precedents.

A hybrid pipeline integrating lexical simplification and syntactic normalization employed by a proposed system. This system uses a combination of techniques to simplify language make the structure of sentences clear and summarize the documents in a way that combines extracting essential information and generating a summary. This is made possible by models like LEGAL-BERT and LEGAL-PEGASUS. To ensure the summaries are accurate and are based on the context; the system uses Retrieval-Augmented Generation to search for information using vectors which enhances the response of the system. This system also has a module to represent entities in a structured way and a layer that allows human beings to check the results to prevent errors and ensure that the system is reliable especially, in important legal cases.

Instead, getting overwhelmed by a 60 - page contract, we get a simple breakdown of what is fair, what is risky and what the fine print actually means. JURISYNTH is a system that provides an easy way to understand and efficiently handle the confidential legal texts and documents to support legal decisions in a reliable way.

Key Words: legal document analysis, legal BERT, legal PEGASUS, Retrieval Augmented Generation, Text Summarization, Legal AI

1. INTRODUCTION

There is a huge increase of legal documents, court hearings, legal decisions and what do judges say; as we all know that legal system is growing digitally faster. The old ways to process this legal work has significantly evolved into inefficiency and are more likely to have mistakes. The consequences are faced by Legal Professionals where they have to go through a lot of literature and source to interpret things and also to find relevant stuff for respective cases.

This complexity and difficulty in understanding terms makes it challenging for people who are not experts leading to a limit for applying information and its usage.

Lately, there has been a lot of breakthroughs and innovations in the field of Artificial Intelligence particularly in Natural Language Processing. It made possible to invent such systems that can address and manage the huge amount of data. The meaning of word, the context and the analyzing power has contributed in excellence for better understanding by the Transformer networks and Transformer – based Models.

Although, even after the progress have made AI Legal Systems enhanced, but there are lot of countermeasures which are still not connected and only focuses on one thing, like summarizing, discovering appropriate information or predicting what will happen; resulting into incapability of legal advice.

To address the issues and rectify the complications, this paper - Jurisynt, an analysis strategic blueprint where the framework that uses Artificial Intelligence to make legal information at more ease of access and efficient in use. The system we proposed integrates with models like Natural Language Processing and retrieval-augmented generation for extracting structured and context-respective insights from complicated legal documents.

The main contributions to this paper are:

- Jurisynt - an analyzing framework which combines summarization, retrieval and prediction into one process.
- The integrated Natural Language Processing system that combines Legal-BERT, Legal-Pegasus and Retrieval-Augmented Generation understanding of the context. for better
- The Jurisynt has the ability to process legal texts as well as speech inputs.
- There are experiments that show the level of consistency, efficiency with the facts of Jurisynt.

2. Literature Survey

Recent advancements in Artificial Intelligence have focused on using special computer models to make legal work easier. These models use datasets and techniques to explain how they make decisions to improve tasks like predicting court judgments summarizing documents and understanding legal reasoning. Despite a lot of progress, the current approaches have limitations when it comes to working with amounts of data combining different systems and being consistent across different legal situations.

[1] Nigam and his team introduced two computer models called NyayaAnumana and INLegalLLaMA which are specifically designed for predicting court judgments in India. These models use datasets and techniques to understand the context and make better predictions. They work well when they are compared to language models but they are mostly used for higher court judgments and are expensive to train and use.

To make the results easier to understand Nigam and his team also proposed two models called [2] TathyaNyaya and FactLegalLLaMA which focus on extracting facts and explaining the predictions. These models make it easier for legal professionals to understand the results by connecting the predictions with the extracted facts. However, the performance of these models depends on the quality of the data preparation and errors in this stage can lead to results.

PredEx [3] is another approach that tries to explain the decisions made by the computer models by using techniques that make the results more understandable to humans. This approach tries to fill the gap between the computer models and the need for transparency in the legal system. While it makes the results more understandable the explanations may not always reflect the reasoning process used by judges especially in complex cases.

Some computer models, like [4] NyayaRAG use a combination of information retrieval and generation to improve the accuracy and reliability of the results. These models first find legal documents and then use them to generate the results. This is especially important in applications, where accuracy and consistency with established precedents are crucial.

In the area of summarizing documents [5] LawSum uses a special approach that minimizes the need for manual annotated datasets. While this makes the process more efficient the generated summaries may not always be consistent. May miss important details.

[6] MILDSum is another approach that focuses on summarizing documents in multiple languages, including Indian languages. This makes it more accessible to a range of users but the performance may not be as good in languages where there is limited training data.

Some researchers, like Shukla [7] and his team have proposed an approach that combines different methods to achieve a balance between accuracy and readability. While this approach works well it can introduce errors. Make the results less accurate.

Kalamkar [8] and his team proposed a way to structure documents into sections like facts, arguments and judgments. This makes it easier to summarize and retrieve the documents. The performance may not be as good if the documents are not well-structured.

Pallavi [9] and her team developed an AI system that combines search and prediction functionalities. While the system is useful it lacks reasoning capabilities and deep integration of modern natural language processing techniques, which limits its effectiveness in complex legal scenarios.

Other researchers, like Verma and Singh [10] have proposed approaches that use mechanisms to make the results more understandable. While these approaches improve transparency they may not always be as accurate as approaches.

Current legal Artificial Intelligence systems mainly focus on tasks like summarization, retrieval and prediction. They do not provide an integrated legal solution. These systems help improve efficiency. However, they struggle to maintain consistency and understand legal language.

Legal documents have terms and contextual meanings. These are hard for AI models to interpret correctly. Another big challenge is balancing explainability with accuracy. Many advanced AI models give the results. They do not clearly explain how they reached those conclusions.

In law transparency and reasoning are crucial. Legal professionals must justify decisions using evidence and legal principles. Legal AI systems also rely heavily on data preparation and retrieval mechanisms. Legal data is usually unstructured.

It needs preprocessing, indexing and organization. If the retrieval process does not provide legal information the AI-generated response may be inaccurate or unreliable.

Despite these challenges Legal Artificial Intelligence has potential. It can simplify work by improving research and reducing manual effort. These systems still need improvements in consistency, contextual understanding, accuracy and explainability. They must become more reliable and effective, in world legal applications.

Table -1: Comparative Analysis Of Existing Approaches

Approach	Methodology	Strengths	Limitations
NyayaAnumana, INLegalLLaM	Transformer based judgment	High Meaning accuracy	Limited scalability
TathyaNyaya	Fact-based explaining AI	Improved transparency	Errorpropagation
PredEx	Explainable ML	Better interpretability	Inconsistent reasoning alignment
NyayaRAG	Retrieval Augmented Generation	Strong contextual grounding	Retrieval dependency
LawSum	Weakly supervised summarization	Scalable, low labeling cost	Inconsistent summaries

Research Gap

From the existing literature, it is observed that current legal AI systems:

a) They Focus on isolated tasks such as summarization, prediction, or retrieval not a complete solutions in.

b) Lack robust contextual grounding across representation, and human-in-the-loop validation which plays a crucial role in the content delivery.

c) This approach aims to provide context-aware, reliable, and explainable legal insights while overcoming the fragmentation and limitations of existing systems which doesn't provide it.

Motivation for Proposed Work

To address these gaps, we propose Jurisynth, a unified legal Document intelligence system that combines Retrieval-Augmented Generation (RAG) with transformer-based summarization to handle complete legal documents more effectively. Existing approaches often struggle with trade-off between explainability and predictive accuracy, are prone to preprocessing errors, and rely heavily on retrieval quality which makes the complete effort of easing the task more complex and fragmented.

Methodology

Jurisynth processes raw legal documents through a structured pipeline to produce context-aware, understandable insights. It integrates preprocessing, embedding, transformer-based output, and explainable AI overview to enable accurate and efficient legal text analysis.

a. System Overview

Jurisynth uses hybrid architecture that combines domain specific NLP models with Retrieval-Augmented Generation (RAG) this helps with general document where looking for similar cases and documents plays an important role, it often falls under the research work but its completely automatic in this project, supporting summarization, precedent retrieval and legal analysis while maintaining contextual consistency.

b. Input Acquisition

The system accepts legal documents in PDF or text format via a web interface. OCR is applied where needed to extract and converted content into the Models Format to process.

c. Preprocessing

The processed text is transformed into semantic vector embedding using domain-adapted language model to support accurate legal similarity retrieval.

d. Embedding and Retrieval

Text is converted into vector embeddings and stored in databases like FAISS or Pinecone. Relevant legal precedents are then retrieved via similarity search to provide contextual grounding.

e. Transformer-Based Legal Analysis

The retrieved context is passed into a large language model adapted for legal analysis through retrieval-augmented contextual grounding. This enables the system to perform tasks such as clause classification, argument mining, legal summarization and contextual legal interpretation.

f. Explainability and Output Generation

The system incorporates explainable AI techniques such as source attribution. Confidence scoring.

g. Query Processing

User queries are interpreted and matched against the indexed legal knowledge base using semantic search. This ensures that the most contextually relevant documents and precedents are surfaced before further analysis is performed.

h. AI Processing

A RAG-based pipeline combines retrieved information with transformer models such as Legal-Pegasus to generate summaries and legal insights. The system also incorporates an explainability component that identifies important reasoning factors.

i. Backend Integration

The backend handles all api based tasks, manages data flow, and formats outputs while ensuring secure and efficient system operation through out the complete process.

j. Output Generation

The system produces structured summaries, key facts, and legal analysis, presented through a user interface and exportable as reports.



Fig -1: Methodology Of Jurisynth

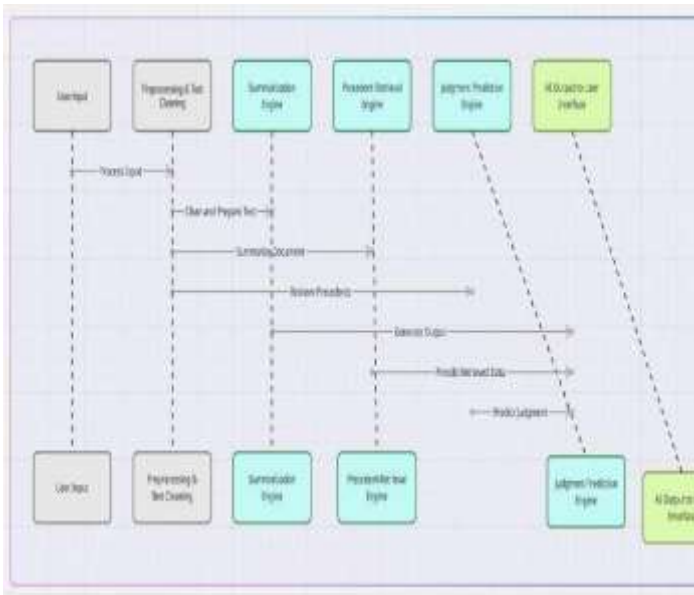


Fig-2: Block Diagram

Fig. 2. Block Diagram illustrates the Jurisynth AI-based legal analysis system. The process begins with user input, followed by preprocessing and embedding generation. Relevant precedents are retrieved from a vector database and combined with the input for transformer-based summarization and prediction. The system refines retrieval if needed and generates structured outputs, which are delivered to the user interface.

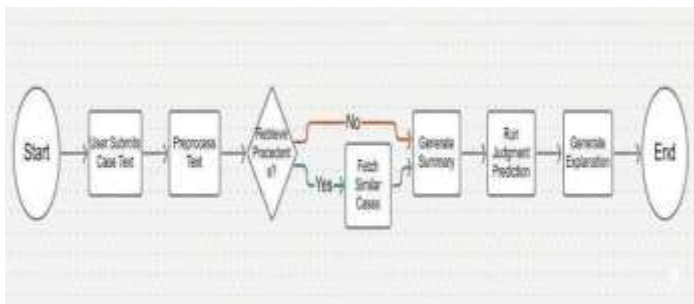


Fig 3.3 -- Flowchart

Fig. 2. Flowchart shows the workflow of the Jurisynth system. The process starts with user input, followed by preprocessing, embedding generation, and retrieval of relevant precedents. The enriched data is processed by transformer-based modules for summarization and prediction. The system refines retrieval if required and generates structured outputs, which are then delivered to the user interface.

3. RESULTS AND EVALUATION

There is testing of the Jurisynth system to evaluate the effectiveness of it. This is accomplished with the use of "appropriate" types of measures, for summarizing and analyzing legal documents. Since there is no set of legal cases to test the Jurisynth system The testing is conducted using actual case studies which are submitted into Jurisynth system The function that the system performs. is checked to see how It is used to address these actual legal cases.

a. Evaluation Metrics

These are the metrics used to measure the effectiveness of system:

- The amount reduced in Document length due to the summarization process: $CR = \text{Summarized document's length} / \text{Original document's length}$ called Compression Ratio (CR).
- Information Coverage (IC) Judges the percentage of key information that is remembered in the created summary: $IC = (\text{Number of Key Facts Extracted} / \text{Total Key Facts in Original}) \times 100$
- A measure of the quality of context preservation. A measure of CPS (Context Preservation Score). $\text{Correct Contextual Statement} / \text{Total Critical Statements}$: How well the system keeps logical and legal relationships between the summarized content.
- Explainability Score (ES) Compares the interpretability of system explanations of their results, $ES = \text{No. of Justified Outputs} / \text{Total Outputs}$

b. Experimental Setup

Assessment is carried out based on an analysis of a sample of documents of law on issues relating to ownership, claims of agreement, possession. The input documents are processed through the Jurisynth pipeline and yield structured data such as summaries, key facts, identified legal issues, and AI-generated insights.

c. Results

System works very well on all assessment measures. An overall effectiveness of the framework is assessed and quantified using a case study.

Metric	Value	Observation
Compression Ratio (CR)	0.34	Significant reduction in document length while preserving meaning
Information Coverage(IC)	86%	Majority of critical legal facts retained
Context Preservation Score (CPS)	0.88	Legal relationships and reasoning maintained accurately
Explainability Score (ES)	0.90	Strong interpretability with clear reasoning provided

Table 2: shows the performance evaluation of the Jurisynth Framework.

a. Analysis

The results indicate that Jurisynth can successfully transform unstructured legal documents into meaningful information, and can also extract important information and compress the text significantly while preserving the important information.

b. Discussion

With retrieval mechanisms for contextual grounding on top, the transformer-based model can attain higher accuracy. To be more convenient and less complicated to use, structured outputs. There are two key steps of fact extraction in Visualization and System Output. These outputs provide evidence of the practicality of the system, finding legal problems and summarizing and analyzing information appropriately.

c. Limitations

Lack of large-scale quantitative evaluation
Dependence on input quality
Limited validation across diverse legal domain.

[7] A. Shukla et al., "Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation," arXiv preprint

[8] Y. Huang, L. Sun, C. Han, and J. Guo, "A High Precision Two-Stage Legal Judgment Summarization," Mathematics, vol. 11, no. 6, p. 1320, 2023.

[9] Y. Pallavi et al., "LEGAL AI: An AI-Powered Legal Research and Case Prediction System for the Indian Judiciary," International Journal of Advanced Research in Computer and Communication Engineering, vol. 14, no. 3, pp. 22–30, 2025.

[10] X. Shang, "A Computational Intelligence Model for Legal Prediction and Decision Support," Computational Intelligence and Neuroscience, vol. 2022, ID 5795189, 2022.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible. So, we acknowledge all those whose guidance and encouragement served as a beacon light and crowned our efforts with success.

With due respect, we thank our Head of Department Dr. Mahesh Wankhade, Computer Science and Engineering, for his motivating support, keen interest which kept our spirits alive all through. We would also like to express our sincere thanks to our Project Coordinator, Prof. Shital Mehta, Department of Computer Science and Engineering, for continuous encouragement, valuable guidance, and support during the development of this project. We would like to express thanks to our guide Prof. Jordan Choudhari Department of Computer Science and Engineering who has guided us throughout the completion of this project.

REFERENCES

[1] S. K. Nigam, B. D. Patnaik, S. Mishra, N. Shallum, K. Ghosh, and A. Bhattacharya, "NyayaAnumana & INLegalLLaMA: The Largest Indian Legal Judgment Prediction Dataset and Specialized Language Model for Enhanced Decision Analysis," arXiv preprint arXiv:2412.08385, 2024.

[2] S. K. Nigam, B. D. Patnaik, S. Mishra, N. Shallum, K. Ghosh, and A. Bhattacharya, "TathyaNyaya and FactLegalLLaMA: Advancing Factual Judgment Prediction and Explanation in the Indian Legal Context," arXiv preprint arXiv:2504.04737, 2025.

[3] S. K. Nigam et al., "PredEx: Legal Judgment Reimagined: PredEx and the Rise of Intelligent AI Interpretation in Indian Courts," arXiv preprint arXiv:2406.04136, 2024.

[4] S. K. Nigam et al., "NyayaRAG: Realistic Legal Judgment Prediction with RAG under the Indian Common Law System," arXiv:2508.00709, 2025. arXiv preprint.

[5] V. Parikh et al., "LawSum: A Weakly Supervised Approach for Indian Legal Document Summarization," arXiv preprint arXiv:2110.01188, 2021.

[6] D. Datta et al., "MILDSum: A Novel Benchmark Dataset for Multilingual Summarization of Indian Legal Case Judgments," arXiv preprint arXiv:2310.18600, 2023.