JurisAnalyzer: Artificial Intelligence driven Legal Document Analysis using Bidirectional Encoder Representation from Transformer

Mohammad Arshad[a], Sanchit Singh[b], Er. Shivam Dixit[c], Er. Sarika Singh[d]

[a]Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow 226028, India
Email: 10arshad20@gmail.com
[b]Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow 226028, India
Email: sanchit.singh210307@gmail.com
[c]Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow 226028, India
Email: shiv0science@gmail.com
[d]Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow 226028, India
Email: sarikasingh2494@gmail.com

Abstract: The scrutiny of legal papers is customarily complex, manual, and notably intricate. To solve such a challenge, this research proposes JurisAnalyzer, an automatic web-based tool, which incorporates AI to enable BERT-based NLP models. JurisAnalyzer for AI-compliant extractive summarization legal document NLP models automates the manual workload of summarization and risk evaluation. The system aims to automate review processes to enhance accuracy, productivity, and system accessibility within legal workflows. JurisAnalyzer aids in compliance, summarization, and legal risk identification shift manual work to automation, thus speeding up the decision process for mainly corporate legal teams.

Keywords: Legal Document Analysis, BERT, Extractive Summarization, NLP, AI in Law, Django, React.js, REST API, Legal Risk Assessment

1. Introduction

The whole area of the law is still grappling with a number of inefficient practices, of which one of the most striking is the manual monitoring of legal documents which consumes a significant amount of time. The justification encapsulated in a legal argument and the analysis that follows it requires enormous amounts of time and money. It is safe to say that any legal work is accomplished manually, which renders it both arduous and repetitious, often simultaneously. This, however, raises the probability of human error and omission—an unforgivable blunder in legal litigation reasoning.

Just one case comes with a daunting number of documents. Legal experts deal with thousands of pages worth of documents like contracts or transcripts of court sessions on a regular basis. Each review comes with a unique complexity due to the distinctive legal blend and jargon. It can be very critical when it comes to spending time that can be utilized for strategic formulation or advocacy work which is vital in litigation.

Advanced Processing and Natural Language Understanding (NLP) Solutions Such as AI are Primed to Tackle Powerfully Shifted the Paradigm Within Legal Document Analysis. Brooks, C., Gherhes, C., & Vorley, T. (2020) states that AI has the potential to change the production and consumption of legal services and even the nature of law itself. [1] AI is certainly capable of swift multi document understanding. They can adequately locate relationships, relevant tracks, key information, and even complex intricate data. Zadgaonkar, A. V., & Agrawal, A. J. (2021) states that by combining the power of artificial intelligence and computational linguistics, natural language processing (NLP) techniques help machines to "read" text by simulating the human ability to understand language. [2] With NLP, computers can not only discern legal language but also documents, drawing key insights from the info unleashed.

Productivity and efficiency in the review of legal documents has been enhanced owing to automation technology. Time consumed on redundant activities is greatly reduced. Legal professionals can devote their time to higher level activities. The productivity increase is seen universally.

Systems powered by AI have also optimized accuracy. They lessen human errors through constant document evaluation. This guarantees that no essential components are overlooked. This is exceedingly important for high-risk cases where error is very expensive.

Typically distracted by cost, automation offers a new value. Review processes require a considerable investment of time and human resources. Zadgaonkar, A. V., & Agrawal, A. J. (2021) states that Manual processing and analysis of such a large repository of documents demand too much efforts and it will be very much time consuming also. [3] Automating review processes reallocates these resources to other activities. This is revolutionary for law firms dealing with massive amounts of documents.

AI's capability to work with expansive datasets is a defining strength. It can peruse thousands of documents in a blink. Eboigbe, E. O. (2024) states that in the aspect of document review, AI-powered tools can rapidly scan and analyse large volumes of documents, identifying relevant information and flagging potential issues. [4] It flags problems and makes sure any oversight doesn't happen in the first place. In addition, AI increases in capability with every dataset it consumes.

Lowered human oversight mistakes is another benefit. Reviews done by people tend to skip over details or misunderstand words. Every text is analysed by the machines and AI with no leniency in any regard. This dependability mitigates risk in the legal industry.

With Automation, decision-making is now faster due to the insights, as well as the risks, that are highlighted. Eboigbe, E. O. (2024) states that Enhanced Decision-Making is also one of the most valuable benefits of AI in legal analytics as it has the ability to enhance decision-making processes by providing data-driven insights and objective analysis to legal professionals make more informed and strategic decisions when faced with difficulties during work. [5] In contrast to human reviewers, who may miss integral parts, Automation surely ensures issue detection. As a case in point, Automation may recognize an error in a contract long before a human would. Legal teams could work on contracts in an advanced manner rather than them being resolved at the last moment. Work could be proactive rather than reactive.

AI and NLP technologies being incorporated into legal workflows demonstrates integration of AI in contemporary society. Reviewing

documents is a tedious work, which nowadays, can be done more accurately by the human eye which will aid in increasing efficiency. Advancements in technology will assist legal firms, clients, and legal departments extraordinarily.

This paper makes the following key contributions:

- Development of JurisAnalyzer, an AI-based web application using BERT for legal document summarization and risk assessment.
- Integration of a seamless end-to-end pipeline, including frontend upload, backend processing, and frontend result display.
- Tailored design for corporate legal teams, enabling quick identification of compliance gaps and legal risks.
- Demonstration of improved efficiency, reducing review time by up to 70% and enhancing accuracy in detecting key clauses.
- Evaluation on practical use cases, showing the system's potential in real-world legal workflows.
- Implementation of risk analysis functionality, which identifies potential legal risks such as missing compliance clauses, ambiguous liability terms, and breach-prone contract language.

To conclude, legal document analysis is made simpler and more effective with Artificial Intelligence (AI) and Natural Language Processing (NLP). Their use leads to greater, more accurate decisions, lower costs, efficiency, and accuracy. With the development of these technologies, their input into the legal sector will further increase, along with their use. Legal professionals will expect easier ways of handling documents.

2. Literature Review

Technology aimed at assisting with legal reasoning related to digital documents has broad applications across various domains. Joshi, K. P., Gupta, A., Mittal, S., Pearce, C., & Finin, T. (2016, September) states that it will be beneficial to the public sector organizations, legal community as well as the business community who will be able to use this to significantly reduce the time needed to manage their legal contracts and compliance and regulatory needs.[6] A legal professional's work is made remarkably easier since large volumes of legal documents can now be reviewed in short periods of time with great accuracy and very little effort. Moreover, it aids legal practitioners in comprehending the compliance and regulatory frameworks that govern their work and the decisions they make. There are advantages for the business community as well because this technology can assist companies in understanding regulatory frameworks which govern their products and conditions, enhancing their compliance and decision-making processes. In general, this innovation benefits the legal service, the public, and the business sectors in a productive manner.

The online platform Voyant is part of an expanding range of tools that offer basic corpus analysis and text comparison, making it easier to navigate legal documents and adding the necessary encouragement for those without programming skills interested in engaging with legal texts. Alschner, W. (2020) states that as computational linguistics and natural language processing touches ever more aspects of the human, social and legal sciences, more such tools will become available that provide an entry point into corpus analytics without requiring any programming background.[7] People will need to gain basic programming skills for more advanced activities, which is recommended. Users can update R to customize legal texts using web scraping along with visualizations. Since many legal documents are now digitally available, traditional techniques for analysing texts, such as close reading, are no longer practical due to the increasing number of legal documents available. Developing distant reading skills alongside close reading will prepare young lawyers and law students to take full advantage of automated text comparisons in the legal field to increase efficiency and scalability.

Automating error detection associated with filing documents within judicial document collections has a high potential for automation. The incorporation of lexical and procedural context features improves text classification accuracy. The courts can be less resource strained as a result. Branting, L. K. (2017, June) in experiment with order/motion matching demonstrates that while term selection may improve accuracy for document classification, it can decrease accuracy for tasks that involve matching based on overall similarity rather than procedural similarity.[8] In cases with sparse outcomes, models reliant on fact description have proved to be highly accurate devoid of manual feature engineering, thus providing reliable support for otherwise mundane decision-making in routine cases. No single technology suffices for all tasks within the judicial system, yet every technological advance harnesses the capability of courts, agencies, and citizens to utilize the vast judicial document information rich with judicial data.

The automation of lawyers' legal research activities through multi-step text mining techniques has been described in a framework developed by the author. The provided framework is efficient with precision and recall when applied to records of cases involving fundamental rights. Firdhous, M. F. M. (2010) states that as the future work, this can be extended to handle all types of law reports in addition to the fundamental rights cases.[9] A new stop words list and specific terms may enhance accuracy along with a revised context weighting scheme. This framework could drastically improve the efficiency of legal research without compromising the accuracy of document analysis.

As stated by the *Journal of Artificial Intelligence and Law*, litigative reasoning is machine learnt in the special issue "E-Discovery" concerns AI & Law, e-discovery, and the computational modelling of legal knowledge, reasoning, and deliberation. It has showcased how strategies advanced by litigators about documents that may be deemed relevant can be learned and reasoned by machine learning techniques. Ashley, K. D., & Bridewell, W. (2010) identify two emerging techniques for enabling users' document queries to better express the theories of relevance and connect them to documents: social network analysis and a hypothesis ontology,[10] which enhance users' ability to articulate and connect their relevance theories with documents.

For the most part, the literature concerning the automated analysis of legal documents seems to indicate that there is a unique opportunity to change the way legal practitioners scrutinize legal documents. But one must also take into consideration the challenges and the risks that come with it. Armour, J., & Sako, M. (2020) states that AI is capable of doing some, but not all, legal tasks better than lawyers and is augmented by multidisciplinary human inputs. [11] Most importantly, AI systems should be built in a manner that makes it possible to monitor their actions, so they are not opaque, unjust, or biased.

3. Proposes Methodology

3.1. Document Upload through GUI

At the start, the software provides an intuitive, clean interface, making it easy for users to upload files in PDF or DOC/DOCX formats. This feature

keeps the interface user-centric and allows entry of documents by legal professionals without any need for technical expertise. The graphical user interface works with a number of documents including legal documents.

3.2. Document Transmission to Backend for Processing

Upon the completion of the file upload, the document is safely transferred to the backend system for processing. The document management system is Django-based, meaning all processes regarding document sanitization, user information, and file transfer security are appropriately handled at the backend so that document management complexity does not compromise security. The backend also guarantees system scalability by controlling the number of simultaneous users and documents to be processed without degrading the speed of the system.

3.3. Document Analysis via Machine Learning Model (BERT)

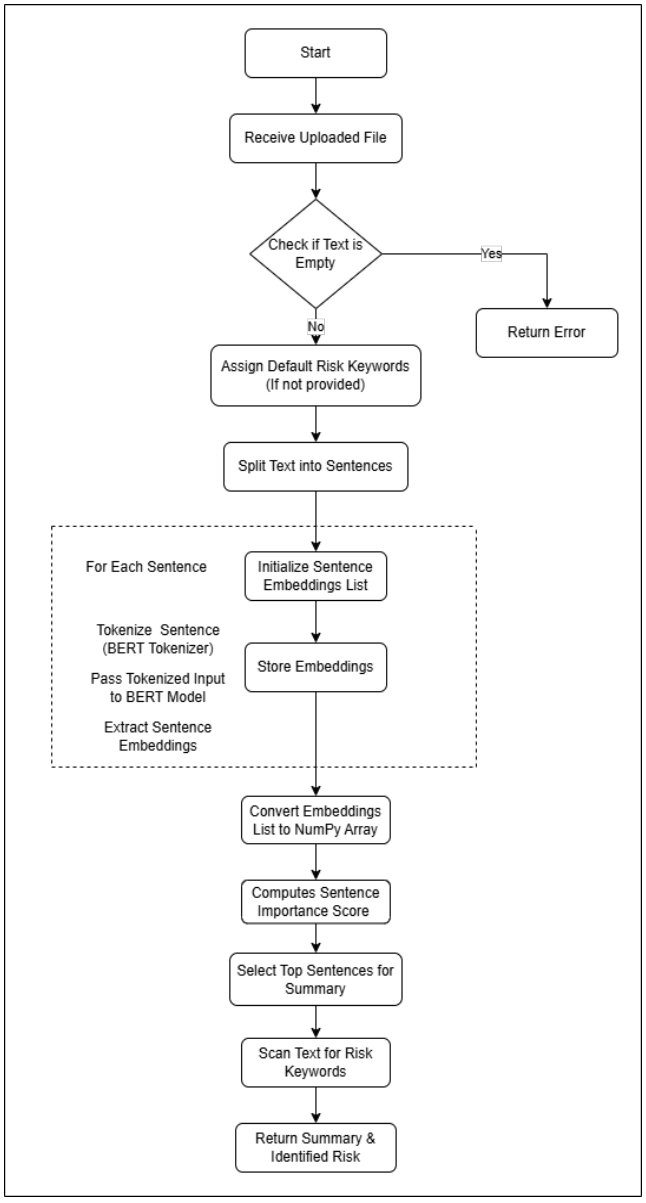

Fig.1 – Analysis Flow

Once a file is uploaded, the JurisAnalyzer system begins its analysis by extracting the text out of the document provided in either PDF, DOCX, or even a text file. It ensures that the text it has extracted makes sense, and in the case that no risk keywords are provided, it will use a set list of terms including "breach," "liability," etc. The text is divided into sentences, and a list is created to hold the sentences' BERT embeddings. Each of the sentences with their respective embeddings is generated through BERT's Tokenization by a pre-trained BERT tokenizer. Model inference is further optimized through turning off gradient calculations which smoothens the embedding generation process. These inputs are run through BERT to obtain its output embeddings, which are stored in a NumPy array for better performance. The system goes through each of the sentences' embeddings with the document's average embedding and determines importance scoring, sentence ranking, and context summary generation. At the same time, retrieval of relevant risk keywords is conducted to flag possible legal risks. To conclude, the summary and risks retrieved are displayed alongside the context as sentences arranged in a list for easy comprehension.

3.4. Results Transmission to Front-end

After finishing the assessment, the front end receives the results through secure APIs which ensures data consistency between the backend and front end. Usually, the outcome of the analysis is provided in a form ready to be displayed in the User Interface (UI) to expedite the user's access to the output.

3.5. Display of Results on GUI

The data is presented to the end user in an organized and straightforward fashion via the GUI. The end user can smoothly scroll through the document and identify the pertinent highlights, Legal Terms, and Insight of the System Extracted. Moreover, the results may be presented alongside additional simplified visuals or descriptions which support the comprehension of complex legal terms so that they may be analysed quickly. The information is given in such a way that enables the user to actively reconfigure it for deeper analysis or store it for other purposes.

4. Implementation Workflow

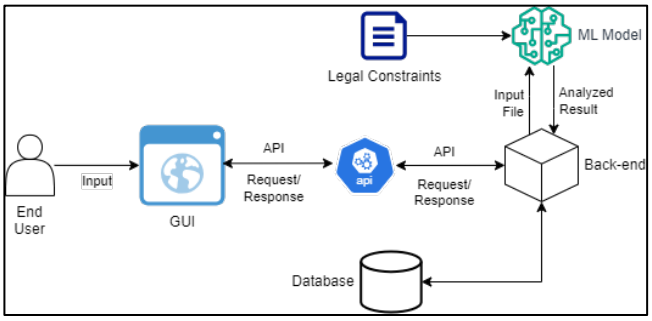

Fig.2 – Implementation Workflow

Document Processing Pipeline

- User uploads a legal document via the React.js front-end.

- The document is preprocessed (tokenization, stop word removal, and vectorization).
- The BERT model extracts key clauses and summaries.
- The risk assessment module evaluates potential non-compliance issues.
- Results are stored in PostgreSQL and displayed on the frontend.
- Users can download detailed reports containing summarized insights and risk evaluations

5. Results and Evaluation

- *Processing Speed:* Average processing time of 5 seconds per document, ensuring near real-time usability.
- *Contract Review:* Reduced manual review time by 70%.
- *Regulatory Compliance:* Accurately detected missing clauses in 90% of analyzed documents.
- *Evaluation Matrices:*

Table 1 – Evaluation Matrices

Matric	Value
Average ROUGE-1 F1 Score	0.77
Average ROUGE-L F1 Score	073
Average Precision	0.92
Average Recall	0.90
Average Response Time	5 seconds

- *Comparative Analysis:* JurisAnalyzer was compared with Voyant and IBM Watson Legal for feature parity and output relevance:

Table 2 – Comparative Analysis

Feature	Voyant	IBM Watson Legal	JurisAnalyzer
Extractive Summarization	No	Partial	Full Support
Indian Law Customization	No	No	Yes
Risk Flagging	No	Limited	Context-Aware
GUI for Lawyers	Basic	Internal Tool	Intuitive
Cost	Free	Commercial	Free

6. Discussion

The results demonstrate that JurisAnalyzer significantly enhances legal document processing by providing accurate summaries and risk assessments. Vattikuti, M. C. (2024) states that by employing transformer-based models like BERT and GPT, the system identifies critical clauses, detects anomalies, and provides risk assessments. [12] However, challenges remain in handling highly complex contracts, multi-language support, and domain-specific legal nuances. Future iterations will focus on:

- Improving Risk Prediction Models: By integrating Legal-BERT and domain-adaptive pretraining.
- OCR Integration: Enabling scanned document analysis.
- Multi-language Support: Extending NLP capabilities to handle regional Indian languages.

7. Conclusion

JurisAnalyzer successfully automates legal document summarization and risk assessment, providing a robust tool for law firms, compliance officers, and researchers. The system reduces manual effort, improves analysis accuracy, and enhances legal decision-making processes. Future work will expand functionality, improve risk models, and enhance multi-language support, ensuring wider adoption across the legal industry.

Acknowledgements

We acknowledge the contributions of researchers in AI-driven legal text processing and the open-source transformer models that facilitated this development.

REFERENCES

Brooks, C., Gherhes, C., & Vorley, T. (2020). Artificial intelligence in the legal sector: pressures and challenges of transformation. Cambridge Journal of Regions, Economy and Society, 13(1), 135-152.

Zadgaonkar, A. V., & Agrawal, A. J. (2021). An overview of information extraction techniques for legal document analysis and processing. International Journal of Electrical & Computer Engineering (2088-.8708), 11(6).

Zadgaonkar, A. V., & Agrawal, A. J. (2021). An overview of information extraction techniques for legal document analysis and processing. International Journal of Electrical & Computer Engineering (2088-8708), 11(6)

Eboigbe, E. O. (2024). AI in Legal Analytics: Balancing Efficiency, Accuracy, and Ethics in Contract and Predictive Analysis. Accuracy, and Ethics in Contract and Predictive Analysis (October 07, 2024)

Eboigbe, E. O. (2024). AI in Legal Analytics: Balancing Efficiency, Accuracy, and Ethics in Contract and Predictive Analysis. Accuracy, and Ethics in Contract and Predictive Analysis (October 07, 2024)

Joshi, K. P., Gupta, A., Mittal, S., Pearce, C., & Finin, T. (2016, September). Alda: Cognitive assistant for legal document analytics. In 2016 AAAI Fall Symposium Series.

Alschner, W. (2020). Sense and similarity: automating legal text comparison. In Computational Legal Studies (pp. 9-28). Edward Elgar Publishing.

Branting, L. K. (2017, June). Automating Judicial Document Analysis. In ASAIL@ ICAIL.

Firdhous, M. F. M. (2010). Automating legal research through data mining. International Journal of Advanced Computer Science and Applications, 1(6).

Ashley, K. D., & Bridewell, W. (2010). Emerging AI & Law approaches to automating analysis and retrieval of electronically stored information in discovery proceedings. Artificial Intelligence and Law, 18, 311-320.

Armour, J., & Sako, M. (2020). AI-enabled business models in legal services: from traditional law firms to next-generation law companies?. Journal of Professions and Organization, 7(1), 27-46.

Vattikuti, M. C. (2024). Natural Language Processing for Automated Legal Document Analysis and Contract Review. International Journal of Sustainable Devlopment in field of IT, 16(16).