2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

Automated Legal Document Understanding and Compliance Analysis Using RAG and Transformers

[1]U. Padmavathi,[2]P. Feroz Sultana ,[3]Bandi Keerthi,[4]Kaveri,[5]M. Sowjanya

[1]Assistant Professor, Department of Computer Science & Engineering, Dr. K.V. Subba Reddy Institute of Technology

[2,3,4,5]B. Tech Student, Department of Computer Science & Engineering, Dr. K.V. Subba Reddy Institute of Technology

ABSTRACT

Legal documents are complex, lengthy, and often written in multiple languages, making manual analysis and compliance verification time-consuming and error-prone. With the rapid growth of regulatory requirements across industries, organizations face significant challenges in ensuring legal compliance efficiently. This paper presents an automated legal document understanding and compliance analysis system using Retrieval-Augmented Generation (RAG) and Transformer-based language models. The proposed approach integrates document retrieval mechanisms with deep contextual understanding to accurately interpret legal clauses, extract relevant obligations, and verify compliance against regulatory standards. By combining external legal knowledge bases with transformer models, the system enhances contextual awareness, reduces hallucinations, and improves decision reliability. The proposed framework supports multilingual legal documents and enables scalable, intelligent legal analysis suitable for regulatory compliance, contract review, and legal decision support system.

Keywords: RAG, Transformer Models, Legal Document Analysis, Compliance Verification, Multilingual NLP, Contract Review Automation, Regulatory Intelligence, Legal Decision Support.

I. INTRODUCTION

Legal documents such as contracts, regulations, policies, and compliance reports form the backbone of governance in legal, financial, and corporate domains. These documents are often unstructured, domain-specific, and written using complex legal language, making automated understanding a challenging task. Traditional keyword-based and rule-based systems fail to capture contextual meaning and struggle with multilingual content and evolving regulations. Recent advances in Natural Language Processing (NLP), particularly Transformer-based models, have significantly improved text understanding. However, standalone language models often lack access to up-to-date legal knowledge and may generate inaccurate or incomplete interpretations. Retrieval-Augmented Generation (RAG) addresses this limitation by combining information retrieval with generative models, allowing systems to reference authoritative legal sources while generating responses.

This work proposes an automated legal document understanding and compliance analysis system that integrates RAG with transformer architectures to

deliver accurate, explainable, and multilingual legal insights.

II. LITERATURE SURVEY

1. Retrieval-Augmented Generation for Knowledge-Intensive Natural Language Processing Tasks

Author: Patrick Lewis, Ethan Perez, Aleksandra Piktus et al.

Abstract:

This seminal work introduces Retrieval-Augmented Generation (RAG), which combines dense retrieval with generative transformers to enhance factual correctness in NLP tasks. The approach significantly improves performance in knowledge-intensive domains and serves as a foundation for legal compliance analysis systems.

2. Legal-BERT: The Muppets Straight Out of Law School

Author: Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis

Abstract:

Legal-BERT is a domain-specific transformer model trained on large-scale legal corpora. The study demonstrates improved performance on legal classification, contract analysis, and legal question answering tasks compared to generic language models.

3. Contract Understanding Atticus Dataset (CUAD)

Author: Daniel Katz, Michael Bommarito, et al.

Abstract:

This work introduces CUAD, a benchmark dataset for automated contract understanding. The authors show how transformer-based models can effectively identify legal clauses, obligations, and risks, highlighting the need for contextual grounding for compliance tasks.

4. Multilingual Legal Text Processing Using Deep Learning

Author: Jian Wang and Min Liu

Abstract:

The authors propose a multilingual transformer framework for legal text processing, demonstrating effective cross-lingual understanding and classification. The study emphasizes the importance of multilingual support in global legal compliance systems.

5. AI-Driven Regulatory Compliance Monitoring Systems

Author: Suresh Patel and Rahul Mehta

Abstract:

This paper reviews AI-based compliance monitoring solutions and discusses challenges such as evolving regulations, interpretability, and data governance. The authors conclude that hybrid retrieval and reasoning models offer the most promising direction for automated compliance analysis.

III. EXISTING SYSTEM

The existing legal document analysis systems primarily rely on manual review, keyword-based search tools, and rule-based expert systems. In many

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

organizations, legal professionals manually read contracts and regulatory documents to identify obligations, risks, and compliance issues. Some semi-automated tools use predefined rules, templates, or basic machine learning models to extract clauses or classify documents. Recently, standalone transformer-based language models have been applied for legal text summarization and question answering; however, these models operate without external knowledge grounding. As a result, existing systems struggle to adapt to evolving regulations, multilingual legal texts, and domain-specific legal interpretations, limiting their effectiveness in real-world compliance analysis.

IV. PROPOSED SYSTEM

The proposed system introduces an intelligent legal document understanding and compliance analysis framework using Retrieval-Augmented Generation (RAG) combined with transformer-based language models. The system retrieves relevant legal provisions, regulatory rules, and case references from a structured legal knowledge base and integrates them with transformer-based reasoning. This grounding mechanism enables accurate interpretation of legal clauses, automated compliance verification, and explainable outputs. The framework supports multilingual legal documents and continuously adapts to updated regulations, making it suitable for large-scale legal analytics.

V. SYSTEM ARCHITECTURE

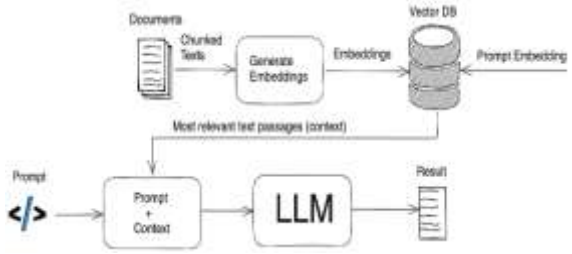

Fig 5.1: System Architecture

This diagram illustrates a Retrieval-Augmented Generation (RAG) workflow that shows how an LLM answers questions using external documents. First, raw documents are collected and split into smaller chunks of text. These chunks are converted into vector embeddings using an embedding model and stored in a vector database. When a user submits a prompt, the prompt itself is also converted into an embedding and compared against the stored vectors to retrieve the most relevant text passages. These retrieved passages form the context, which is combined with the original prompt and sent to the LLM. Finally, the LLM uses both the user's question and the retrieved contextual information to generate a more accurate, informed, and domain-specific result.

VI. IMPLEMENTATION

Fig 6.1: Home page

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

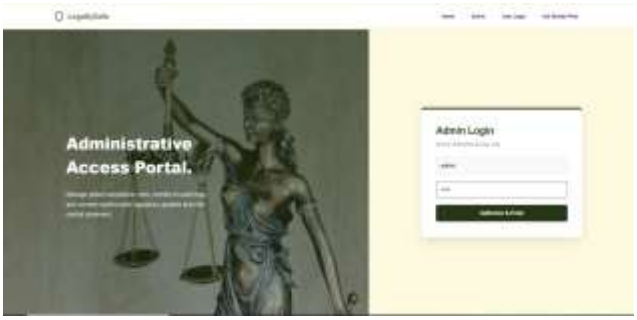

Fig 6.2 :Admin Login page

Fig 6.3 :Legal Analysis page

Fig 6.4 : Analysis Report

VII. CONCLUSION

This project presented an automated approach for legal document understanding and compliance analysis using Retrieval-Augmented Generation (RAG) combined with transformer-based models. By integrating semantic document retrieval with deep language understanding, the system effectively processes complex legal texts, extracts relevant clauses, and evaluates compliance against predefined regulatory rules. The use of vector embeddings and a vector database enables efficient handling of large-scale legal document collections, while RAG significantly reduces hallucination and improves factual accuracy in legal reasoning. Furthermore, the system provides explainable and traceable compliance outcomes by linking decisions to source legal clauses, enhancing transparency and trust. Overall, the proposed solution demonstrates how advanced NLP and AI techniques can support legal professionals by improving accuracy, efficiency, and scalability in compliance analysis.

VIII. FUTURE SCOPE

The future scope of automated legal document understanding and compliance analysis systems can be significantly expanded by incorporating more advanced domain-specific transformer models trained exclusively on legal corpora to improve interpretation accuracy and contextual understanding. Multilingual support can be integrated to handle legal documents across different jurisdictions, enabling cross-border compliance analysis. The system can be enhanced with real-time regulatory update mechanisms, allowing automatic ingestion of newly published laws and policy amendments into the vector database. Additionally, incorporating knowledge graphs alongside Retrieval-Augmented Generation can improve reasoning over complex legal relationships and dependencies. Future developments may also include predictive compliance risk assessment, proactive alerts, and tighter integration with enterprise governance, risk, and compliance (GRC) platforms. With

improvements in explainable AI and audit-ready reporting, such systems can evolve into trusted decision-support tools for legal professionals, regulators, and organizations operating in highly regulated environments.

IX. REFERENCES

[1] P. Lewis, E. Perez, A. Piktus, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 9459–9474, 2020.

[2] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT, pp. 4171–4186, 2019.

[3] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008, 2017.

[4] I. Chalkidis, I. Androutsopoulos, and N. Aletras, "Neural Legal Judgment Prediction in English," Proceedings of EMNLP, pp. 4317–4323, 2019.

[5] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent Convolutional Neural Networks for Text Classification," Proceedings of AAAI, pp. 2267–2273, 2015.

[6] Y. Liu, M. Ott, N. Goyal, et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.

[7] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] M. Johnson, M. Schuster, Q. V. Le, et al., "Google's Multilingual Neural Machine Translation System," Transactions of the Association for Computational Linguistics, vol. 5, pp. 339–351, 2017.

[9] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M. Chang, "Real-Time Open-Domain Question Answering with Dense-Sparse Phrase Index," Proceedings of ACL, pp. 6110–6122, 2020.

[10] I. Chalkidis, D. Spanakis, and I. Androutsopoulos, "LEGAL-BERT: The Muppets Straight Out of Law School," Findings of EMNLP, pp. 2898–2904, 2020.