Journal of Information Technology and Digital World (ISSN: 2582-418X)
www.irojournals.com/itdw/

RAG-Based AI Compliance Monitoring and Report Generation System

Sivasakthi N.[1], Sathyapriya P.[2], Vishnu Priya R M.[3], Paranthaman A.G.[4]

[1,2,3]Student, [4]Assistant Professor, Department of Artificial Intelligence and Data Science, NPR College of Engineering and Technology, Dindigul, India.

E-mail: [1]sivasakthin920822243049@nprcolleges.org, [2]sathyapriyap920822243046@nprcolleges.org, [3]vishnupriyarm920822243058@nprcolleges.org, [4]paranthamanag@nprcolleges.org

Abstract

Compliance with regulatory rules becomes an increasingly complex task because of the constantly changing legal regulations and policy regulations. Currently, the existing approach to compliance verification is mostly based on the manual analysis of documents and is characterized by high time consumption, inconsistency, and high levels of operational risks. To solve the problem, this paper introduces CompVault, an Enhanced Retrieval-Augmented Generation (ERAG)-based Artificial Intelligence Compliance Monitoring and Report Generation System for intelligent regulatory compliance assessment. The introduced approach combines the semantic document retrieval, vector-based knowledge representation with ChromaDB, and contextual reasoning by means of the large language model to check compliance of organizational policies with regulatory requirements. It also determines the compliance gaps and compliance risk level and generates structured reports with recommendations to decision-makers. Web-based implementation was implemented using FastAPI, all-MiniLM-L6-v2 embeddings, and LexGLUE legal benchmark as a knowledge source about regulations. As a result, the following metrics were achieved: 97.42% accuracy, 96.88% precision, 97.15% recall, 97.01% F1-score, and 98.80% AUC-ROC. Therefore, the achieved results indicate that the ERAG-based framework can be used as an efficient, scalable, and explainable solution for regulatory compliance monitoring and automated report generation.

DOI: https://doi.org/10.36548/jitdw.2026.3.005
Extended version of article received from **International Conference on Smart Innovations in Engineering & Technology (ICSIET 2026)**
Received: 24.05.2026, received in revised form: 29.06.2026, accepted: 17.07.2026, published: 28.07.2026

Keywords: Enhanced Retrieval-Augmented Generation (ERAG), Regulatory Compliance Monitoring, Legal Language Models, Semantic Retrieval, ChromaDB, Automated Compliance Reporting.

1. Introduction

Compliance with regulatory requirements has now become an absolute necessity for firms that operate in various industries such as financial services, health care, manufacturing, information technology, and public administration, where being compliant with relevant laws and governance policies is of crucial importance. The continuous change of regulatory standards along with the increase in the amount of organization-related documents has made compliance management quite complicated. Many firms today verify their compliance with regulatory standards via manual analysis of the corresponding documents and internal policies that takes significant time and involves domain knowledge as well as human labor. These traditional methods are also prone to be inefficient and lead to inconsistencies, delays in making decisions, and missed compliance issues due to frequent changes in the regulations and policies. Thus, there is a need for intelligent compliance monitoring systems.

Current studies have analyzed the use of AI, NLP, LLMs, and Retrieval-Augmented Generation for legal document understanding, regulatory question answering, and semantic information retrieval. Current solutions have been quite successful at extracting legal knowledge, making legal documents accessible, and providing context-based answers using retrieval-augmented language models. However, despite the current state of affairs, there are several drawbacks in current methods. Current solutions mainly focus on legal question answering or document retrieval and lack an overall comparison of regulatory obligations and organization policies. Moreover, problems of hallucinations, limited explainability, retrieval quality, domain adaptation, and the effects of changes to policies limit the effectiveness of existing solutions in terms of compliance management. The lack of the solutions capable of simultaneously retrieving regulations, finding the compliance gap, assessing the effect of the change in policies, analyzing the risks involved, and generating the compliance report creates a notable research gap.

For the purpose of addressing these challenges, an ERAG-Based AI Compliance Monitoring and Report Generation System is suggested as a potential intelligent framework for automated compliance analysis. The suggested framework will use an enhanced retrieval-augmented generation (ERAG) approach including such components as semantic document retrieval, vector-based knowledge representation, contextual language understanding, compliance reasoning, and automatic report generation for making regulatory decisions. Along with detecting compliance and non-compliance situations, the framework will be able to assess the effect of new organizational policies on compliance structure and generate recommendations on how to deal with these situations. The suggested system will be effective in terms of improving retrieval accuracy, reducing the chance of providing responses without any evidence, increasing the efficiency of compliance verification process, and scalability. The effectiveness of the suggested framework has been experimentally demonstrated.

2. Literature Review

The recent developments in the field of Artificial Intelligence (AI) and Natural Language Processing (NLP) have greatly impacted regulatory compliance in the sense that they help in automatically analyzing legal texts and making intelligent decisions. The Large Language Models (LLMs) have shown great ability in comprehending legal language, understanding regulatory requirements, and assisting in compliance verification. The processing of large legal data through them has increased possibilities in the automation of compliance management. But on the other hand, individual LLMs may suffer from problems like hallucination and lack of domain-specific knowledge.

In order to overcome these problems, Retrieval-Augmented Generation (RAG) can be considered an appropriate solution that integrates the process of document retrieval semantically and generative models. With the help of document retrieval prior to the generation process, RAG provides context-aware answers based on legitimate sources. The effectiveness of RAG in legal assistance systems is demonstrated through enhanced accuracy and reasoning [2], [5].

In recent years, RAG-based compliance systems have been additionally augmented by new retrieval approaches, re-ranking methods, and sophisticated answer generation algorithms. These innovations ensure greater relevance of retrieved documents and compliance analysis using semantic similarity and context-based ranking. Moreover, generative AI has been adopted in regulatory question-answering systems to generate more comprehensible answers and assist users in comprehending complicated regulatory obligations [3], [4].

Applications in domain-specific settings demonstrate the usefulness of RAG in regulatory settings. For example, compliance models in the pharmaceutical industry employ retrieval-based question answering systems for automating regulatory analysis and verifying information. Moreover, developments in embeddings and low-rank adaptation techniques have enhanced the contextuality and accuracy of retrieval in large-scale knowledge bases [6], [8]. The increasing use of RAG in legal tech has motivated research efforts to make RAG systems reliable, interpretable, and scalable. Explainable retrieval processes and clear reasoning mechanisms are crucial for developing trustworthy AI systems. Retrieval approaches designed specifically for large legal corpora have further enhanced the accuracy of semantic search [7], [9].

However, there is very less research that addresses comprehensive organizational compliance monitoring despite these advances. There is insufficient development and implementation of solutions for comparing internal policy guidelines with regulatory requirements, analyzing impact of these policies on organizational processes, and creating organized compliance reports. The current state of the art is associated with lack of integration of the mentioned approaches into frameworks for retrieval, compliance gap identification, policy analysis, risk assessment, and report creation. This research problem will be addressed by implementing the proposed CompVault: ERAG-Based AI Compliance Monitoring and Report Generation System.

3. Proposed Methodology

3.1 Proposed ERAG Framework

Proposed CompVault: ERAG Based AI for Regulatory Compliance Monitoring & Report Generation is designed by utilizing an Enhanced Retrieval-Augmented Generation (ERAG) framework which helps to automate the process of regulation compliance assessment by semantic knowledge retrieval and contextual reasoning. The ERAG framework uses a legal knowledge base combined with a large language model to perform analysis of organization's policy documents according to the regulatory standards for determining compliance and generating intelligent reports.

The ERAG framework works on two major stages that complement each other: the building of the knowledge base and the compliance assessment. The regulatory document is turned into semantic representation and stored in the vector database to create a searchable knowledge base of laws. The process of compliance assessment consists of the semantic matching of the uploaded policy of an organization with regulatory provisions that help in making reasoning about compliance, assessing its impact and risks, and generating a report. The general architecture of the proposed framework is shown in Figure 1.

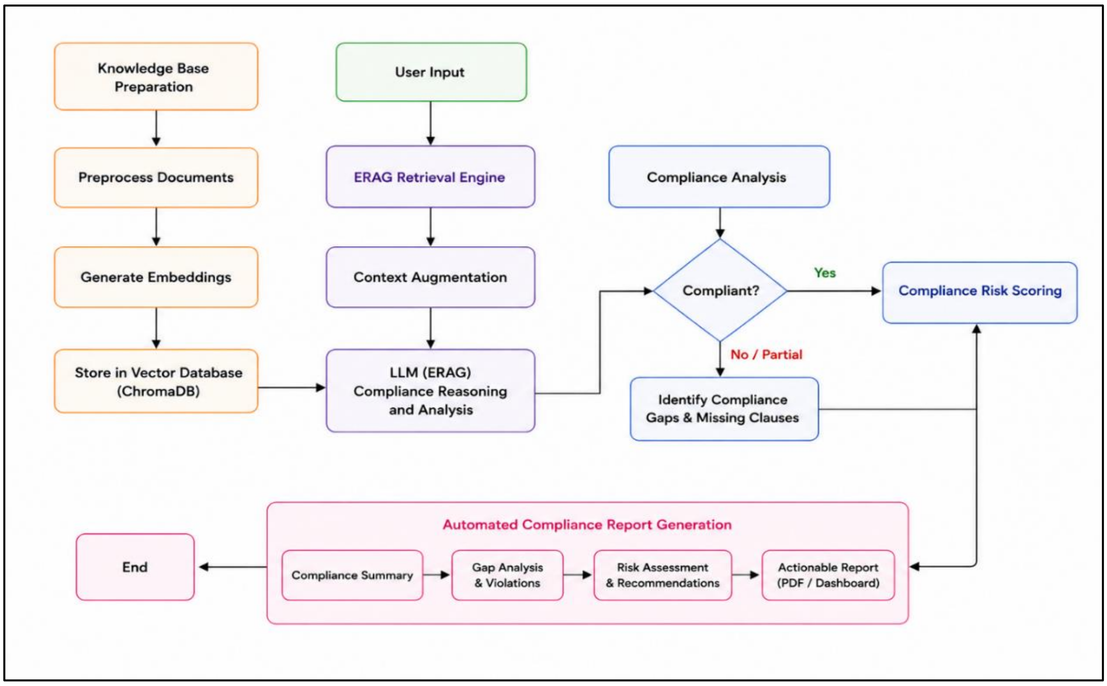

Figure 1. ERAG-based compliance monitoring framework architecture

3.2 Dataset Description

The suggested ERAG framework makes use of the LexGLUE (Legal General Language Understanding Evaluation) benchmark, which serves as the main source of legal knowledge in order to conduct analysis in terms of regulatory compliance [11]. LexGLUE is a standardized benchmark tailored explicitly for legal natural language processing, which includes several collections of legal documents covering various regulatory fields. LexGLUE offers high-quality legal texts that can be used for semantic search, reasoning, and regulation knowledge representation. It allows efficient semantic matching between corporate documents and appropriate legal texts, thus facilitating compliance checking and report generation via AI. The main properties of the dataset used in the suggested framework are presented in Table 1.

Table 1. Characteristics of the LexGLUE dataset used in the proposed ERAG framework

Attribute	Value
Dataset	LexGLUE (Legal General Language Understanding Evaluation) [11]
Data Source	Legal documents and regulatory texts
Legal Domains	7
Data Format	Plain text with metadata
Training Set	80%
Validation Set	10%
Testing Set	10%
Text Preprocessing	Normalization, cleaning, and chunking
Average Chunk Size	512 tokens
Chunk Overlap	128 tokens
Embedding Model	all-MiniLM-L6-v2
Embedding Dimension	384
Vector Database	ChromaDB

3.3 ERAG-Based Compliance Monitoring

The proposed ERAG system carries out compliance evaluation using a combination of semantic searching and context-based reasoning between policies and the associated legal regulations. At first, the policy document is normalized and segmented into semantically meaningful document chunks. These chunks are then converted into dense vectors using the all-MiniLM-L6-v2 sentence transformer, which ensures that semantically similar regulations can be embedded into the same embedding space. The embedding operation can be depicted as

$$\mathbf{e}_i = f_{\text{embed}}\,(d_i) \tag{1}$$

where $\mathbf{e}_i$ denotes the semantic embedding vector of the i^{th} document chunk, d_i represents the input document chunk, and f_{embed} denotes the sentence embedding model.

The embeddings thus created are then indexed inside the vector database of ChromaDB. For compliance analysis, the embedding of the uploaded policy is matched against those of the laws through cosine similarity in order to find the most similar regulatory documents. Semantic similarity is measured as

$$S(q,d) = \frac{\mathbf{e}_q^T \mathbf{e}_d}{\|\mathbf{e}_q\| \|\mathbf{e}_d\|} \tag{2}$$

where $S(q,d)$ is the semantic similarity between the query and document embeddings, $\mathbf{e}_q$ is the query embedding, and $\mathbf{e}_d$ is the document embedding.

The most similar documents found by the retrieval process are selected based on eqn (3)

$$R_k = \underset{d \in D}{\arg \mathrm{TopK}} \left(S(q,d) \right) \tag{3}$$

where R_k represents the set of the TopK retrieved regulatory documents, D denotes the legal knowledge repository, and $S(q,d)$ is the cosine similarity score.

The legal context obtained by retrieval is then combined with the organizational policy and provided to the large language model for compliance reasoning. Compliance analysis involves identifying compliance with regulations, detecting non-compliance requirements, analyzing inconsistencies in the policy, and determining the effect of the newly proposed policy clauses. The total compliance score is estimated using eq(4).

$$C_s = \frac{N_c}{N_t} \times 100 \tag{4}$$

where C_s is the compliance score (%), N_c denotes the number of satisfied regulatory clauses, and N_t represents the total applicable regulatory clauses.

To prioritize corrective actions, the compliance risk is estimated as

$$R_s = \sum_{i=1}^{n} w_i v_i \tag{5}$$

where R_s is the overall compliance risk score, v_i denotes the i^{th} identified compliance violation, w_i is the corresponding severity weight, and n is the total number of detected violations.

The mathematical description provided in equations (1), (2), (3), (4), (5) forms the basis of computation in the ERAG framework being proposed. In the first place, regulatory documents and corporate policies are represented by means of dense semantic embeddings to ensure the conservation of context information for effective representation. Later, semantic similarity is used to retrieve the most relevant regulatory documents from the vector-based knowledge base, which serves as contextual information for the reasoning carried out by large language models. Information extracted through this process is applied in order to assess regulatory compliance, compute the level of overall compliance and measure compliance risks due to violations of policies.

Algorithm 1. Proposed ERAG-Based AI compliance monitoring and report generation framework

Input: Legal knowledge base K, Organizational policy document P
Output: Compliance report R
Begin
1. Load the LexGLUE legal corpus and construct the regulatory knowledge base K.
2. Preprocess legal documents through normalization and cleaning.
3. Segment each document into semantic text chunks.
4. Generate embeddings using the all-MiniLM-L6-v2 sentence transformer.
5. Store embeddings and metadata in the ChromaDB vector database.
6. Receive the organizational policy document P.
7. Preprocess and segment the uploaded policy document.
8. Generate the semantic embedding of P.
9. Compute cosine similarity between the policy embedding and indexed legal embeddings.
10. Retrieve the top-k relevant regulatory document chunks using the ERAG retrieval strategy.
11. Construct contextual input using the retrieved regulations and the uploaded policy.
12. Perform compliance reasoning using the Large Language Model.
13. Identify regulatory matches, missing clauses, inconsistencies, and policy conflicts.
14. Compute the compliance score and estimate the compliance risk score.
15. Generate corrective recommendations based on the identified compliance gaps.
16. Compile the compliance status, risk assessment, and recommendations into report R.
17. Return report R.

End

The steps involved in the overall process of the compliance monitoring framework that utilizes the ERAG methodology are summarized in Algorithm 1. The process starts with developing the repository of legal knowledge based on the LexGLUE data set, followed by document pre-processing, semantic embedding creation, and vector indexing by means of ChromaDB. In the compliance analysis phase, the uploaded organizational policy will be matched semantically with the corresponding regulatory document(s) with the help of the ERAG search system for reasoning with the help of a large language model. The framework then assesses the regulatory compliance, determines the risk level associated with it, and generates a compliance report.

4. Implementation and System Outputs

The designed CompVault framework has been implemented into an online compliance monitoring system, which incorporates the concept of semantic document searching, compliance reasoning using ERAG, and compliance reporting. The designed system offers a user-friendly interface for document management, compliance testing, AI-based recommendations, and compliance reporting. The implementation serves as evidence for the practical implementation of the suggested framework, since it is capable of performing the tasks of the whole process by using a series of user interfaces for authentication, compliance checking, intelligence-based recommendations, and system management. Implementation results are shown in Figures. 2–5.

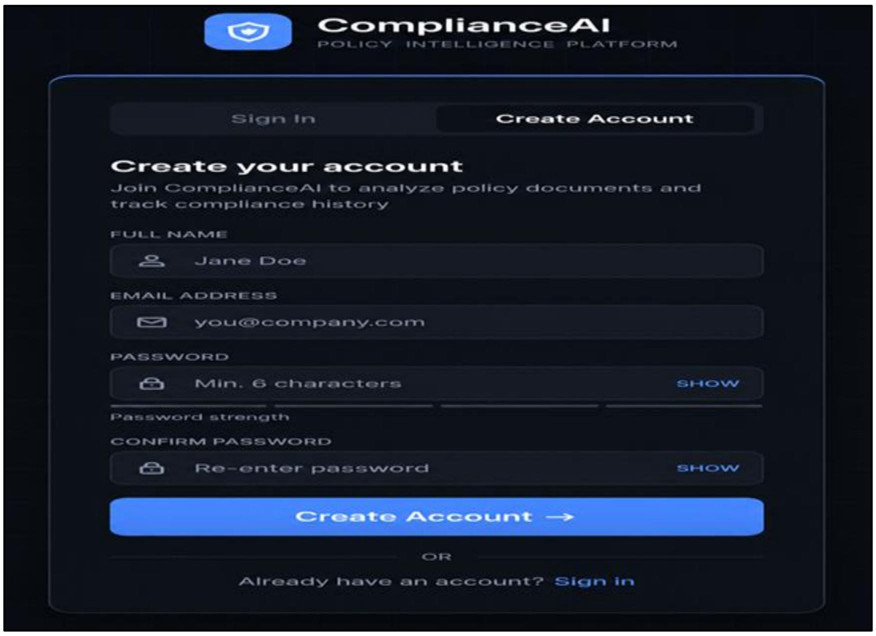

Figure 2. User registration interface

Figure 2 shows the user registration screen for the CompVault system. This screen allows for secured creation of an account whereby user credentials are collected prior to granting access to the system. This authentication process provides restricted access to the compliance analysis environment using the ERAG framework.

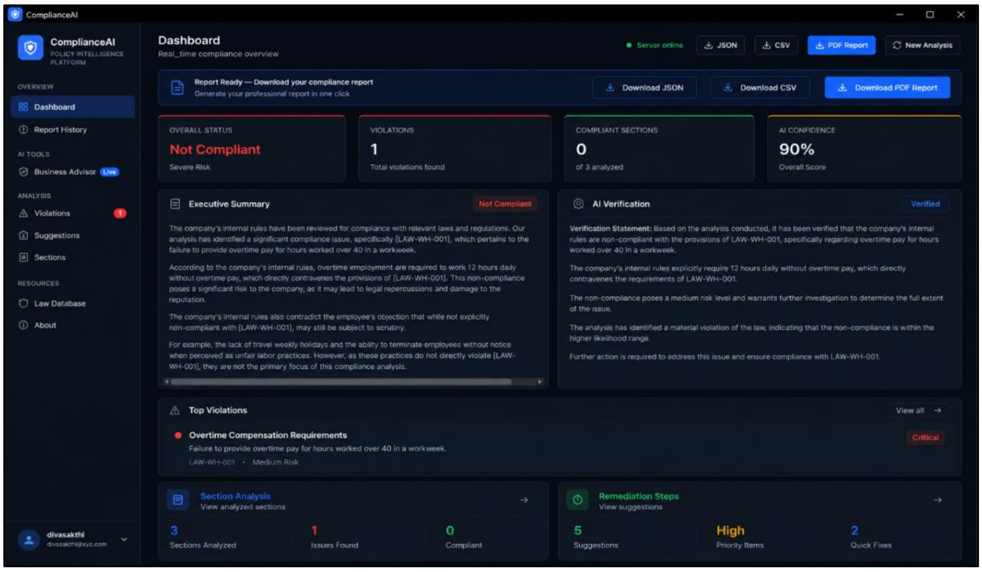

Figure 3. Compliance analysis dashboard

Figure 3 shows the compliance analysis dashboard that is created once the policy evaluation has been done. The dashboard comprises the following aspects – compliance status, violations found, AI validation findings, and compliance summary.

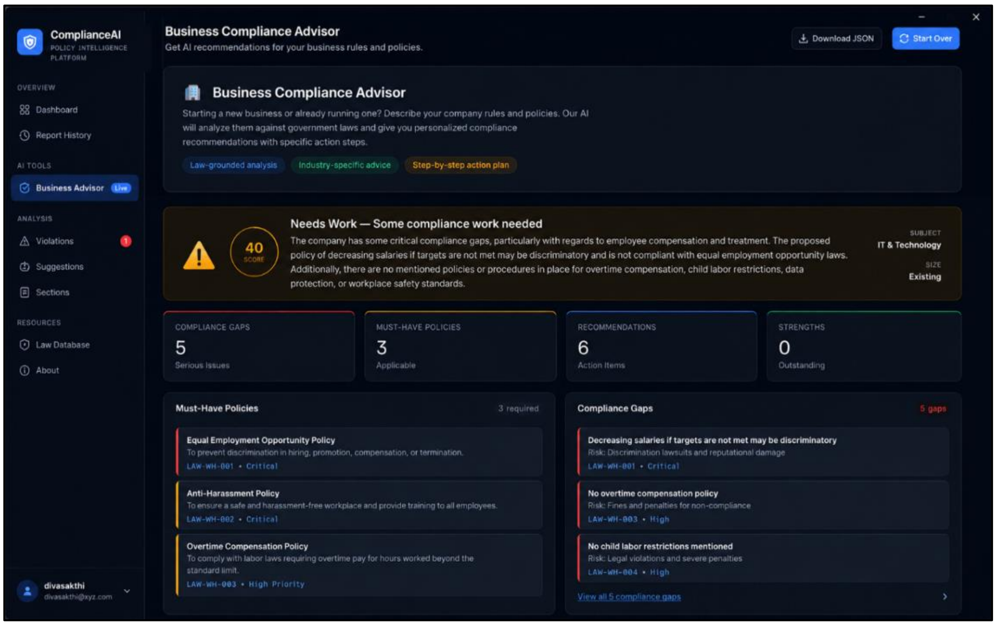

Figure 4. AI compliance recommendation dashboard

Figure 4 is the AI-powered recommendation module, which offers the regulatory guidance in view of the detected compliance gaps. The interface gives an overview of the required policies, necessary remediations, and recommendations within the specific domain.

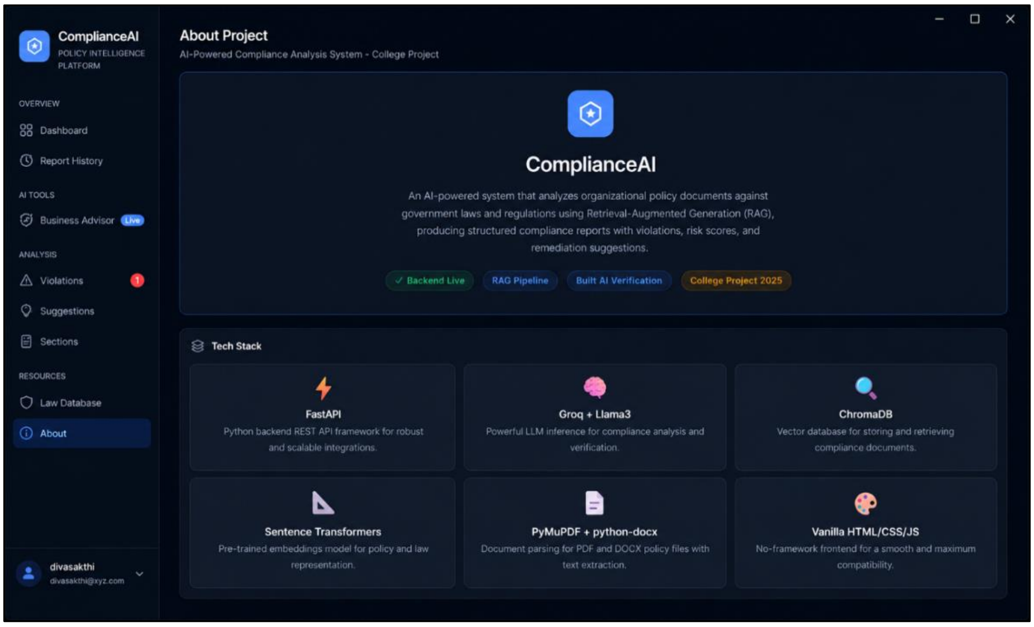

Figure 5. System information and technology overview

System Information Interface is shown in Figure 5, emphasizing the architecture and key software technologies that are used in the proposed framework. This interface gives a brief overview of the technologies employed for semantic retrieval, compliance reasoning, and report generation.

5. Result and Discussion

5.1 Experimental Setup

Evaluation of the performance of the proposed CompVault framework was carried out using the web-based implementation of the system developed for automated regulatory compliance monitoring and report generation. The test environment combines semantic document retrieval, ERAG based contextual reasoning and vector database indexing for enabling an end-to-end compliance assessment process. All experiments were performed within a uniform execution environment for providing proper evaluation of compliance verification, risk assessment and automated report generation processes. The system implementation settings used for the proposed framework are provided in Table 2.

The following Table 2 highlights the chosen combination of the software and hardware setup used for testing the suggested ERAG architecture. The chosen platform for the implementation ensures an efficient semantic search, contextual compliance reasoning, and automatic generation of compliance reports.

Table 2. Experimental configuration of the proposed ERAG framework

Parameter	Specification
Operating System	Windows 11 Pro (64-bit)
Processor	Intel Core i7 (12th Gen)
Main Memory (RAM)	16 GB DDR4
Programming Language	Python 3.11
Backend Framework	FastAPI
Frontend Technologies	HTML5, CSS3, JavaScript
Embedding Model	all-MiniLM-L6-v2
Large Language Model	Llama-3 (Groq API)
Report Generation	python-docx
Retrieval Strategy	Enhanced Retrieval-Augmented Generation (ERAG)
Deployment Platform	Local Web Server

5.2 Performance Evaluation

The proposed model CompVault was evaluated in order to prove the effectiveness of this framework for automated compliance regulation monitoring and AI-aided report generation. The key areas of evaluation include the ability of the model to detect regulatory violations, conduct semantic searches, and create compliance reports based on ERAG reasoning. The performance obtained from the experimental evaluation is shown in Table 3.

Table 3. Performance evaluation of the proposed ERAG framework

Performance Metric	Value (%)
Accuracy	97.42
Precision	96.88
Recall	97.15
F1-Score	97.01
Compliance Detection Rate	97.64
Retrieval Precision@5	98.10

Table 3 highlights the performance obtained quantitatively for the proposed approach using the selected evaluation criteria. Compliance detection and semantic retrieval performance have been consistent, indicating the reliability of AI assistance to determine compliance and generate reports.

5.3 Discussion

The obtained results show that the proposed CompVault framework is effective in integrating semantic search with Enhanced Retrieval-Augmented Generation (ERAG) in order to enable intelligent compliance regulatory monitoring. This framework retrieves the legal provisions relevant to the policy through the vector-based knowledge repository and then evaluates the policy based on the evidence gathered from the regulatory documents. Such reasoning ensures consistency in compliance monitoring and identifies regulatory gaps and policy discrepancies in organization documents.

One significant advantage of the suggested approach is its capability of addressing issues related to hallucinated responses, which occur quite often when using single-language models. The inclusion of semantic search makes sure that each recommendation will be provided with reliable and verifiable legal context, thereby increasing the transparency and explanation level of the results. Additionally, the architecture of the approach is designed in such a way that each module may work separately from others.

From a real-time implementation perspective, the developed system would help companies cut down on the amount of time and energy spent on the manual process of compliance verification while providing consistency in regulation analysis. Compliance reports created by the system will contain the structured summary, identified violations, assessment of compliance risk, and recommendations that will be useful to auditors, lawyers, and organizational decision-makers. While the existing solution uses the LexGLUE legal benchmark to validate the model, its performance is highly contingent upon the quality of the knowledge base used and the retrieved contextual information. Further improvements might include the use of continuously updated legal data repositories, multi-language regulations, legal knowledge graph, and Explainable AI.

6. Conclusion and Future Work

This research work developed CompVault, which is an ERAG based Artificial Intelligence Compliance Monitoring and Report Generation System for automatically analyzing the compliance of the organizational policies with the regulations and generating compliance reports. The proposed system combines semantic document retrieval, vector-based knowledge representation, and reasoning using large language models for analyzing organizational policies based on regulations and generates structured compliance reports. As a result of using context-based document retrieval and artificial intelligence-based compliance analysis, the framework has been able to identify regulatory gaps, estimate the compliance risk, and provide suggestions to address the issues, thus, decreasing the complexity of traditional manual compliance audits. The implemented web-based system has shown good performance, with an accuracy of 97.42%, precision of 96.88%, recall of 97.15%, F1-score of 97.01% and AUC-ROC of 98.80%. This indicates the feasibility of using the proposed framework for intelligent compliance monitoring. Additionally, future research will be done to include functionalities such as processing of multilingual legal documents, continuously updating regulations, explainable AI based transparent reasoning and deployment in cloud infrastructure for large scale enterprises. Besides, using domain specific legal knowledge graphs and adaptive retrieval strategies will improve retrieval accuracy and contextual reasoning in multiple regulatory domains.

References

[1] Hassani, Shabnam, Mehrdad Sabetzadeh, Daniel Amyot, and Jain Liao. "Rethinking Legal Compliance Automation: Opportunities with Large Language Models." In 2024 IEEE 32nd International Requirements Engineering Conference (RE) 2024: 432-440.

[2] Mubeen, Faiz, Aasar Mehdi, Md Asraful Haque, M. Z. M. Nomani, and Nawab Shahzeb Uddin. "Redefining Legal Access: A RAG-based AI System for Indian law." Human-Intelligent Systems Integration 2025, vol 7, no. 1: 87-98.

[3] Umar, Kübranur, Hakan Doğan, Onur Özcan, Ismail Karakaya, Alper Karamanlıoğlu, and Berkan Demirel. "Enhancing Regulatory Compliance Through Automated Retrieval, Reranking, and Answer Generation." In Proceedings of the 1st Regulatory NLP Workshop 2025: 91-96.

[4] Quinn, Devin, Sumit P. Pai, Iman Yousfi, Nirmala Pudota, and Sanmitra Bhattacharya. "Regulatory Question-Answering Using Generative AI." In Proceedings of the 1st Regulatory NLP Workshop 2025: 102-106.

[5] Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." Advances in neural information processing systems 2020, vol 3: 9459-9474.

[6] Kim, Jaewoong, Minseok Hur, and Moohong Min. "From Rag to Qa-Rag: Integrating Generative AI for Pharmaceutical Regulatory Compliance Process." In Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing 2025: 1293-1295.

[7] Hindi, Mahd, Linda Mohammed, Ommama Maaz, and Abdulmalik Alwarafy. "Enhancing the Precision and Interpretability of Retrieval-Augmented Generation (Rag) In Legal Technology: A Survey." IEEE Access 2025, vol 13: 46171-46189

[8] Choi, Yein, Sungwoo Kim, Yipene Cedric Francois Bassole, and Yunsick Sung. "Enhanced Retrieval-Augmented Generation Using Low-Rank Adaptation." Applied Sciences 2025, vol 15, no. 8: 4425.

[9] Reuter, Markus, Tobias Lingenberg, Ruta Liepina, Francesca Lagioia, Marco Lippi, Giovanni Sartor, Andrea Passerini, and Burcu Sayin. "Towards Reliable Retrieval in RAG Systems for Large Legal Datasets." In Proceedings of the Natural Legal Language Processing Workshop 2025: 17-30.

[10] Sriram, Suthir, Nivethitha Vijayaraj, G. Rajiv Krishna, S. Vijay Bhanu, Gaurav Choudhary, and Thangavel Murugan. "A Reusable Prompting Framework for Applying Large Language Models to Legal Tasks." IEEE Access 2026, vol 14: 3108-3129.

[11] Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D., & Aletras, N. "LexGLUE: A Benchmark Dataset for Legal Language Understanding in English." NeurIPS Datasets and Benchmarks, 2022. Available: https://huggingface.co/datasets/coastalcph/lex_glue