An Intelligent NABL Laboratory Information System Using Retrieval-Augmented Generation For Compliance Automation

Shivani S. Shelar [1], Sandeep R. Jadhav[2]
[2]AssistantProfessor
[1, 2] Shree Mahaveer Education Society's Sanghavi College of Engineering, Nashik

Abstract- *Laboratories accredited by the National Accreditation Board for Testing and Calibration Laboratories (NABL) handle large amounts of documentation relating to quality manuals, audit trail records, standard operating procedures, and compliance documents. Traditional systems for managing documents, using keywords for search and manual navigation, have created inefficiencies during audits or when verifying compliance.[4][5] As a result, the development of AI technologies, including Retrieval-Augmented Generation (RAG), indicates the potential for intelligent ways to retrieve documents and obtain contextually relevant answers to questions. In this paper, we will outline our vision for a Laboratory Information System (LIS) fueled by artificial intelligence (AI) through the integration of RAG with the current NABL accreditation workflow. Our proposed LIS system combines semantic document retrieval from vector databases with fact-based and context-sensitive responses generated by large-language models. [1][3]. To this end, we emphasize the creation of a layered architecture based on four components: Document Preprocessing; Vector Embedding; Similarity Searches; and Natural-Language Generation (NLG) Components. Our approach significantly enhances the way laboratories audit their operations, reduces the time required to search for documents, and increases the ability to provide traceable responses generated by the AI component. Our methodology provides laboratories with the opportunity to integrate RAG technology into their operations while building a framework for building trusted AI-based accreditation systems.[6],[10]*

Keywords- Retrieval-Augmented Generation, NABL, Laboratory Information System, Compliance Automation, Semantic Search, Artificial Intelligence

I. INTRODUCTION

An essential role of the National Accreditation Board for Testing and Calibration Laboratories (NABL) is to ensure the quality and technical competence of testing laboratories in addition to providing a source of internationally accepted results for testing laboratories in India. NABL functions under the Quality Council of India (QCI) and follows international standards, including ISO/IEC 17025 and ISO 15189, when issuing and maintaining accreditations[5].Obtaining and retaining NABL accreditation for labs involves extensive amounts of documented evidence being maintained (e.g., Quality Manual, Calibration Records, Audit Reports, SOPs, etc.), so the processes involved in managing, maintaining and retrieving the pertinent information from the various large documentation repositories can be very time consuming and difficult to oversee due to the amount of documentation, and as such, the volume of information obtained through the documentation process can lead to errors in documentation (especially during audits and compliance assessments). The systems that are presently available for Laboratory Information Systems (LIS) provide significant benefits with respect to the handling of information (storage and workflow), however, they are Consulting and analyzing documentation is now significantly easier due to advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP), as well as through the development of new approaches to information retrieval, which combines the best aspects of retrieval models and generative models, known as Retrieval-Augmented Generation (RAG). As a result of RAG, it is now possible for systems to retrieve relevant documents from extremely large document stores and produce responses to user queries generated based on factual content found within thosedocuments; the utilizing of RAG is particularly useful in regulated environments where I. Accuracy II. Transparency III. Traceability is mandatory. This research has proposed a new Intelligent NABL Laboratory Information System, incorporating RAG for the purposes of providing Semantic Document Retrieval, AI Supported Compliance and Readiness for Audit purposes[3][10].

1.1 IMPORTANCE OF AUDIT PREPAREDNESS

NABL audits require traceable documentation for every test and calibration performed. Missing or delayed information can cause non-conformities, affecting laboratory credibility. RAG can help by linking audit queries directly to relevant documents and past audit records.

1.2 CHALLENGES IN TRADITIONAL LABORATORY INFORMATION SYSTEM

- Manual metadata tagging is inconsistent.
- Keyword searches fail for synonyms (e.g., "calibration" vs. "verification").
- Difficult to retrieve historical corrective actions or linked SOPs.

1.3 BENEFITS OF INTELLIGENT LABORATORY INFORMATION SYSTEM

Contextual retrieval can reduce human error, improve compliance, and allow auditors and lab personnel to quickly access evidence without exhaustive manual searches.

1.4 MOTIVATION

In times of increasing digitalization, accredited laboratories by NABL continue to face problems in managing and retrieving compliance information. During internal audits, surveillance assessments, and renewals of accreditation, laboratory personnel must quickly access specific clauses related to historical audit findings, corrective actions, and calibration evidence.

The existing LIS platforms do not have any semantic understanding and also do not support natural language queries or any contextual reasoning. Manual document searches are time-consuming, error-prone, and highly dependent on user expertise. Conventional systems also cannot provide responses that are explainable or traceable to NABL clauses.

There is a distinct need for an intelligent information system that will enable semantic document retrieval, allow for natural language interaction, and ensure responses based on NABL authoritative documentation. Retrieval-Augmented Generation overcomes these shortcomings through the integration of retrieval and generation mechanisms, making it eminently suitable for compliance-driven laboratory environments.

II. LITERATURE REVIEW

The RAG (Retrieval-Augmented Generations) framework was introduced by Lewis et al. (2020) as an effective approach to addressing knowledge-intensive Natural Language Processing tasks. The RAG combines dense retrieval mechanisms and generative models to improve the factual accuracy and decrease the number of hallucinations within the content produced by a Language Model.

In their study, Saleem (2021) provided an overview of the digital transformation challenges faced by NABL accredited laboratories and indicated areas for improvement due to the inefficiencies associated with manually handling documents and preparing for audits. Although the author provides a suggestion for the use of Cloud-Based Laboratory Information Management Systems (LIMS), there was no mention of the use of Artificial Intelligence (AI) and/or semantic retrieval solutions.[1]

Denneman(2023) studied enterprise-level RAG architectures and emphasized the importance of domain-specific fine-tuning of generative models and the development of efficient retrieval pipelines using vector databases such as FAISS. The author supports the use of RAG as an effective approach within specialized or niche domains, including Legal Documentation and Compliance Documentation.

IBM Research (2024) studied the potential use of RAG in developing trustworthy AI systems for use in Regulated Industries. The authors highlight the key benefits of an RAG Architecture, including the ability to trace the citations and generate an audit trail. [3],[10]

Previous studies on RAG mainly focus on general enterprise knowledge systems, while laboratory accreditation presents a unique high-stakes, compliance-intensive domain [1], [3].

Existing LIS solutions concentrate on workflow automation but lack semantic reasoning and contextual retrieval capabilities [4], [7].

Trustworthy AI research highlights that regulated sectors require explainable AI by default, not as an optional feature [3], [10].

Current cloud-based LIMS platforms emphasize scalability but provide limited support for semantic compliance verification [6].

The Quality Council of India (2023) calls for the digitization and automation of the workflows associated with NABL accreditations; however, they did not discuss any use of AI-based retrieval systems or intelligent compliance systems.

A.SEMANTIC EMBEDDINGS FOR DOCUMENT RETRIEVAL

Word embeddings map textual data into dense vector representations that preserve semantic relationships. Mikolov et al. introduced Word2Vec, capturing word-level semantic similarity. BERT, proposed by Devlin et al., provides contextual embeddings improving NLP tasks. Sentence-BERT optimizes transformer models for sentence-level semantic similarity, crucial for retrieving compliance documents in NABL laboratories.

Model	Description	Application in LIS
Word2Vec	Word-level embeddings	Quick keyword expansion
BERT	Contextualized word embeddings	Clause-level understanding
Sentence-BERT	Sentence-level embeddings	Semantic document retrieval

Table 1: Examples of semantic embeddings in LIS

B. VECTOR DATABASES AND SIMILARITY SEARCH

Vector databases like FAISS allow high-dimensional embeddings to be stored and queried efficiently. They support approximate nearest neighbor (ANN) search, reducing query latency for large document corpora. In NABL LIS, vector databases enable

- Quick retrieval of audit clauses.
- Similarity-based document ranking.
- Integration with RAG pipelines for context-aware responses.

C. RETRIEVAL AUGMENTED GENERATION MODELS

RAG combines retrieval and generation:

1. Retrieve relevant documents using semantic embeddings.
2. Use generative LLMs to produce context-aware responses.

Studies like Lewis et al. demonstrate that RAG significantly reduces hallucinations compared to pure LLM responses, a key requirement for compliance systems. REALM and other retrieval-augmented models validate knowledge-intensive NLP applications.

D.LARGE LANGUAGE MODELS IN INFORMATION SYSTEMS

LLMs like GPT-3 and GPT-4 show few-shot learning abilities, enabling understanding of complex compliance questions. Coupled with retrieval, they can:

- Summarize audit reports.
- Answer clause-specific queries.

E. RAG-BASED SYSTEMS IN REGULATORY DOMAINS

Applications in healthcare, legal, and finance illustrate RAG's capability to improve compliance verification. However, limited research targets laboratory accreditation systems, leaving a gap for NABL-specific RAG implementations.

RESEARCH GAPS IDENTIFIED

- No Domain Specific RAG Models Existed for NABL Accreditation Data
- Current LIS Platforms Have Limited Automation and Semantic Understanding
- Current AI Document Systems Do Not Provide Sufficient Explainability and Traceability
- Integrated Framework Between RAG Architecture and LIS Does Not Exist
- Insufficient Evaluation of RAG Performance with Compliance-heavy Industries
- Provide traceable, evidence-based explanations.

III. PROBLEM STATEMENT AND GOALS

3.1 IDENTIFICATION OF ISSUES ASSOCIATED WITH THE PROBLEM

Laboratories accredited by NABL maintain vast amounts of both structured and unstructured documentation. Without a smart electronic document retrieval system that can retrieve all types of documentation in one place, compliance audits may be inefficient, compliance risks will rise due to a lack of documentation, and laboratory operations could experience unnecessary delays caused by the absence of an "intelligent" document retrieval system that can interpret natural language search requests and return context-based responses.[3][5]

3.2 OBJECTIVES

- To analyze the applicability of Retrieval-Augmented Generation for NABL laboratory data.
- To design an intelligent document retrieval system integrated with LIS.
- To enhance audit readiness through AI-assisted, contextual information retrieval.
- To identify limitations and propose improvements for RAG deployment in compliance automation.

IV. PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY

The Modular 5-Layer Architecture Proposed for NABL LIS System:

4.1 SYSTEM ARCHITECTURE

1. Data Source Layer - NABL Accreditation Manuals; SOPs; Audit Records; Quality Documents

2. Data Processing Layer - Include Extracting Text; Cleaning; Chunking; Metadata Tagging

3. Vector Database Layer - Store the Semantic Embedding of a Document to Perform Similarity Search with FAISS

4. RAG Processing Layer - Using LangChain and LLMs Combine Retrieval & Generation Functions

5. User Interface Layer - Upload Documents for Users to Interact with the System

Through this architecture, the system supports Modularity, Scalability, and Accurate AI-Generated
Retrievals.[6][8]

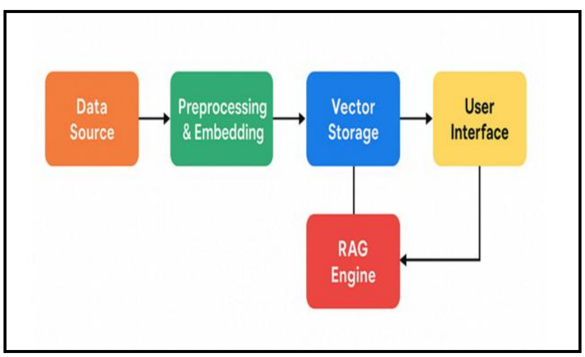

Figure.4.1 Block Diagram of System Architecture

The block diagram illustrates a Retrieval-Augmented Generation-based information system architecture devised to support intelligent, context-aware information retrieval and response generation from large document repositories.

a. Data Source
Overview:

The Data Source block represents the set of raw documents that the system works with. The documents can be in the form of manuals, reports, guidelines, policies, standards, PDFs, text files, or database records.

Functioning:

- Documents are collected from the concerned authorities
- Data can be structured or in unstructured format
- This layer underpins the knowledge base

b. Preprocessing & Embedding

Description:

This block will be responsible for converting raw documents into a machine-readable format with the help of NLP techniques.

Functioning:

- Text is extracted from documents.
- Operations include noise removal, normalization, and tokenization.
- Large documents are divided into smaller, logically coherent pieces
- The pre-trained embeddings models such as Sentence-BERT, OpenAI embeddings convert each chunk to a dense numerical vector or embedding.
- The embeddings capture the semantic meaning of the text

c. Vector Storage

Description:

The Vector Storage block provides a vector database that can store document embeddings and enable efficient similarity searches.

Functioning:

- The generated embeddings are indexed and stored.
- Similarity search algorithms are utilized, such as cosine similarity.
- Enables the retrieval of semantically similar documents efficiently.

- Some common implementations are FAISS, Pinecone, or Chroma.

d. RAG Engine

Description:

RAG Engine is the central intelligent component that does document retrieval and then integrates it with response generation from large language models.

Functioning:

- Receives query embeddings from the user interface
- Retrieves the most relevant document vectors from vector storage
- Expands the user's query with the document context retrieved.
- Passes the augmented context to a Large Language Model (LLM) This produces a factually grounded and context-aware response.

e. User Interface Description
:

The User Interface is where users type in requests or queries and receive answers interactively. Functioning: - Takes natural language input from users - Sends queries to the RAG Engine - Shows generated responses and their respective document references Facilitates ease of access to accurate and explainable information.

4.2 RETRIEVAL AUGMENTED GENERATION (RAG) ALGORITHM

RAG is an AI hybrid framework that incorporates neural information retrieval into large language models to make their responses truly grounded in external knowledge sources. Unlike purely generative models, RAG retrieves relevant documents from a knowledge base to augment response generation with the external information therein, reducing hallucinations and improving explainability.

A. ALGORITHMIC WORKFLOW

Step 1: Document Gathering and Preprocessing

Documents related to NABL include ISO/IEC 17025 standards, quality manuals, SOPs, audit reports, and calibration records, which are collected in their digital formats. The preprocessing consists of text extraction, removal of noise, normalization, tokenization, and chunking.

Step 2: Text Embedding Generation

Every document chunk is then represented as a dense vector using pre-trained embedding models such as Sentence-BERT or OpenAI embeddings.

Step 3: Vector database indexing

These generated embeddings are stored in a vector database like FAISS to enable similarity-based retrievals efficiently.

Step 4: Query Processing

User queries are preprocessed and embedded through the same embedding model to ensure consistency in a semantic space.

Step 5: Semantic Retrieval

The query embedding retrieves the top-k most relevant document chunks by cosine similarity.

Step 6: Context Augmentation

The retrieved documents are joined as an augmented prompt with the query.

Step 7: Response Generation

The large language model produces a response that is grounded and coherent and follows the standards of NABL.

Step 8: Output Presentation

The final response is provided together with extracts from the supporting documents and clause references.

4.3 METHODOLOGY

The methodology consists of six phases:

Phase 1: The researchers collected all NABL documents and prepared them for analysis.

Phase 2:The team used either the OpenAI or Sentence-BERT models to create "embeddings" for the NABL document data.

Phase 3: The researchers created an RAG pipeline to generate responses to the NABL documents based on the provided context.

Phase 4: The researchers developed and implemented a design for the final product.

Phase 5: The researchers conducted evaluations of their final product against three criteria; accuracy, relevance, and response time.

Phase 6: The researchers optimized their final product for performance and assessed its scalability. [3],[9]

V. MATHEMATICAL MODEL

The System is defined mathematically as follows:

S = { I, P, R, O, F } [1][2]

Definition of each component:
I = Input: documents and user queries.
P = Preprocessing.
R = Retrieval based on vector similarity.
O = The output response.
F = The generation function using an LLM.
Using cosine similarity returns the top-k most relevant document chunks for factual and contextually accurate responses.[9],[10]

VI.SYSTEM IMPLEMENTATION FRAMEWORK

The proposed RAG-enabled Laboratory Information System (LIS) can be implemented using a modular microservices-based architecture to ensure scalability and maintainability. Each module is independently deployable, allowing laboratories of different sizes to customize the system according to their operational needs. The system is deployed in a hybrid cloud environment to ensure both data security and computationalefficiency.The ingestion pipeline begins with automated document loaders that extract content from PDF files, scanned audit reports, spreadsheets, and database exports. Optical Character Recognition (OCR) is applied where required to convert scanned documents into machine-readable text. A preprocessing engine cleans the data by removing irrelevant headers, footers, and repetitive boilerplate content. Logical chunking is applied based on document structure such as clauses, sub-clauses, and procedural steps[1][6]

The embedding service uses transformer-based sentence embedding models to ensure semantic consistency. The vector database layer is implemented using FAISS for on-premise deployment or Pinecone for cloud-based scalability. This ensures low-latency similarity search even with millions of document vectors. The RAG engine orchestrates retrieval and generation using LangChain pipelines that integrate document retrieval, prompt construction, and response generation through large language models.

Role-based access control (RBAC) is implemented to ensure that auditors, quality managers, and laboratory technicians only access relevant data. Audit logs are maintained for every query and response, supporting traceability and accountability during regulatory inspections. [1][6][8][9]

The modular RAG architecture allows independent upgrading of retrieval and generation components, reducing system downtime and improving maintainability in regulated environments [1], [6].

Use of vector databases enables laboratories to scale from thousands to millions of documents without degrading retrieval accuracy [6], [9].Integration of LangChain pipelines ensures standardized orchestration of document loaders, retrievers, and LLMs, reducing implementation complexity [8].

Enterprise-grade RAG systems enable hybrid deployment models (on-premise + cloud), which is essential for NABL laboratories with strict data governance requirements [10].

VII. EXPECTED OUTCOMES

- Reduction in document retrieval time.
- Improved contextual accuracy during NABL audits.
- Transparent, source-linked responses.
- Enhanced compliance verification and decision support.

Feature	**Keyword-based Labaoratory Information System**	**RAG-based Laboratory Information System**
Retrieval speed	Medium	High
Contextual understanding	Low	High
Explainable reasoning	Low	High
Audit preparedness	Medium	High

Table 2: Comparison: Keyword Labaoratory Information System vs RAG Labaoratory Information System

VIII. ANTICIPATED OUTCOMES & EXAMINATION

It is anticipated that the new RAG constructed LIS will achieve the following objectives:

1.Significantly shorten document search times versus keyword-based search methods.
2.Improve accurate and relevant response rates during audits.
3.Provide increased transparency by linking the AI provided response directly to the original source document.
4. Enable rapid compliance verification and decision-making.
The new model differs from existing traditional LIS solutions by providing semantic comprehension, contextual reasoning and audit trails - which are essential for compliance with NABL standards.[1][6]

IX. .SECURITY, PRIVACY, AND ETHICAL CONSIDERATIONS

Security and privacy are critical in compliance-driven environments such as NABL-accredited laboratories. The proposed system follows a privacy-by-design approach. All sensitive laboratory data, including patient-related test results and calibration histories, are encrypted at rest and in transit using industry-standard encryption protocols. [3]

Access control policies are enforced through multi-factor authentication and role-based authorization. The system also supports on-premise deployment for laboratories that cannot store sensitive data on public clouds due to regulatory constraints. Data anonymization techniques are applied when using documents for model testing and evaluation.

From an ethical perspective, the system ensures that AI-generated responses remain transparent and explainable. Every generated response is accompanied by source citations, allowing users to verify the authenticity of the information. This mitigates the risk of blind reliance on AI outputs and aligns with NABL's emphasis on accountability and traceability.[5]

Bias mitigation is addressed by continuously updating the document repository and embedding models to reflect the latest NABL guidelines and standards. Human-in-the-loop mechanisms are incorporated so that domain experts can validate critical compliance recommendations.[3][5][9]

Trustworthy AI frameworks emphasize that traceability of AI decisions is mandatory in regulated industries such as laboratory accreditation [3], [10].

NABL guidelines stress the importance of document version control and audit trails, which can be automated using AI-assisted LIS platforms [5].

RAG-based systems support explainability-by-design, as every answer is derived from verified documentation instead of model memory alone [3].

Ethical AI adoption in laboratories requires human oversight for critical compliance decisions, ensuring accountability remains with certified professionals [7].

X. CONCLUSION

Retrieval-Augmented Generation was shown to have great potential to impact how NABL laboratories complete workflows due to the transformation of the workflow through use of an integrated Laboratory Information System (LIS). By building upon the LIS-integration framework, the new framework could drastically improve accuracy, transparency and audit readiness for laboratories. With this framework, we fill critical gaps in current LIS products and create a foundation for trustworthy artificial intelligence in aspirations to achieve accreditation. The focus of future work will include NABL API/ Web Service Integration in live-time, Multilingual Query Support for all languages, development of Explainable AI Framework(s) for NABL laboratories, and a large scale performance evaluation(s) of the system on actual audits datasets.[3][9]

REFERENCES

[1] P. Lewis, E. Perez, A. Piktus, et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, 2020, pp. 1–12.
[2] F. Denneman, "Deep dive into RAG architecture," *LinkedIn Research Notes*, 2023. [Online]. Available: https://www.linkedin.com
[3] IBM Research, "Retrieval-augmented generation for trustworthy AI," *IBM AI Blog*, 2024. [Online]. Available: https://www.ibm.com
[4] M. Saleem, "Understanding NABL audits," *Journal of Quality Systems*, vol. 13, no. 2, pp. 45–52, 2021.
[5] Quality Council of India (QCI), *NABL accreditation guidelines*, New Delhi, India, QCI Report, 2023.
[6] Google Cloud, "RAG implementation strategies," *Google Cloud AI Documentation*, 2024. [Online]. Available: https://cloud.google.com
[7] A. Tolk, "AI chatbots in laboratory systems," *Frontiers in Artificial Intelligence*, vol. 6, pp. 1–10, 2023.

[8] LangChain, “Building domain-specific RAG applications,” *LangChain Documentation*, 2024. [Online]. Available: https://www.langchain.com
[9] NVIDIA Research, “RAG for enterprise knowledge systems,” *NVIDIA AI Journal*, 2025.
[10] IBM, “Architectures for generative AI and retrieval-augmented generation,” *IBM Journal of Research and Development,* vol. 68, no. 1, pp. 1–14, 2024.