AI-Powered Compliance Assistant (FINCOMPLY)

Capstone Project

Phase – I Report

Submitted by

22BCE10681 Giriraj Parsewar

22BCE11351 Anjasi Jaiman

22BCE10847 Riya Patel

22BCE11410 Isha Rani

22BCE10007 Avyaan Verma

in partial fulfillment of the requirements for the degree of

Bachelor of Engineering and Technology

VIT®
BHOPAL
www.vitbhopal.ac.in

VIT Bhopal University
Bhopal
Madhya Pradesh
April, 2025

Bonafide Certificate

Certified that this project report titled “AI Powered Compliance Assistant (FINCOMPLY)” is the bonafide work of “22BCE10681 Giriraj Parsewar, 22BCE11351 Anjasi Jaiman, 22BCE10847 Riya Patel, 22BCE10410 Isha Rani, 22BCE10007 Avyaan Verma” who carried out the project work under my supervision.

This project report (Phase I) is submitted for the Project Viva-Voce examination held on 18 September 2026

Supervisor-
Mr Vishal Singh Bhatti

Comments & Signature (Reviewer 1)

Comments & Signature (Reviewer 2)

Declaration of Originality

We, hereby declare that this report entitled AI Powered Compliance Assistant (FINCOMPLY) represents our original work carried out for the Capstone project as a student of VIT Bhopal University and, to the best of our knowledge, it contains no material previously published or written by another person, nor any material presented for the award of any other degree or diploma of VIT Bhopal University or any other institution. Works of other authors cited in this report have been duly acknowledged under the section "References".

Date- 16 April 2025

Reg No & Name

1. 22BCE10681 Giriraj Parsewar
2. 22BCE11351 Anjasi Jaiman
3. 22BCE10847 Riya Patel
4. 22BCE10410 Isha Rani
5. 22BCE10007 Avyaan Verma

Acknowledgement

This work has benefited in various ways from several people. Whilst it would be simple to name them all, it would not be easy to thank them enough.

We would first like to express our deepest gratitude to our supervisor, Mr. Vishal Singh Bhatti, for his unwavering support, valuable insights, and consistent encouragement throughout the course of this project. His expertise and guidance were instrumental in shaping our ideas and refining our approach at every stage.

We are also thankful to our faculty members and lab staff who provided the necessary technical assistance and infrastructure for the successful execution of our work.

We would like to extend our sincere appreciation to our peers and teammates for their collaboration, dedication, and collective effort, which made this project a truly enriching experience.

Last but not least, we are grateful to our families and friends for their constant motivation, understanding, and encouragement during the entire project timeline.

Together, we have learned, grown, and accomplished something we are truly proud of.

Abstract

Regulatory compliance is one of the biggest challenges faced by companies in India, especially with the frequent updates issued by the Securities and Exchange Board of India (SEBI). Companies often struggle to identify which regulations apply to them and how to implement these guidelines correctly. This leads to compliance risks, penalties, and loss of investor trust.

Our project, FinComply, is an AI-powered solution that uses Large Language Models (LLMs) with Retrieval-Augmented Generation (RAG) to make SEBI compliance easier, faster, and more reliable. FinComply automatically retrieves SEBI circulars and guidelines, summarizes them, and tailors the output to the specific profile of a company (e.g., listed/unlisted status, sector, and size). It also ensures transparency by providing source citations and continuous monitoring of new SEBI updates.

By combining natural language processing, regulatory knowledge, and explainable AI, FinComply aims to reduce compliance burden, minimize risks of violations, and improve trust between companies, regulators, and investors.

CHAPTER-1

INTRODUCTION

The financial market in India is regulated by the **Securities and Exchange Board of India (SEBI)**, which ensures transparency, fairness, and investor protection. However, SEBI issues frequent circulars, notifications, and guidelines that companies must follow to remain compliant. For businesses, especially listed companies, keeping track of these regulations and applying them correctly is a major challenge.

Non-compliance can lead to **heavy penalties, reputational loss, and even restrictions on market activity**. Therefore, there is a strong need for an intelligent system that can **simplify compliance, personalize regulatory information, and ensure traceable adherence** to SEBI guidelines.

Our project, **FinComply**, aims to address this problem by leveraging **Large Language Models (LLMs) with Retrieval-Augmented Generation (RAG)** to create a smart compliance assistant.

1.1 Problem Identification

Fintech organizations deploy microservices rapidly to meet market demands. However, every release must comply with complex regulatory frameworks such as Anti-Money Laundering (AML), Know Your Customer (KYC), PCI-DSS, GDPR, and internal security policies. Currently, compliance checks are often carried out manually, which leads to:

1. **Slow and inconsistent validation** of microservices and infrastructure.
2. **Increased regulatory risk**, as misconfigurations or non-compliance may go undetected.
3. **Slower release cycles**, since compliance reviews delay CI/CD pipelines.
4. **Vulnerable production environments**, caused by human oversight in enforcing policies.

This gap between rapid software delivery and manual compliance processes makes organizations susceptible to legal penalties, reputational damage, and security breaches.

1.2 Applications

FinComply is designed to be an **AI-powered compliance assistant** that simplifies how companies interact with SEBI regulations. The application will cover multiple use cases across compliance, auditing, and monitoring.

Core Applications of FinComply

1. **Automated SEBI Document Summarization**
    - Input: Raw SEBI circulars, notifications, or guidelines.
    - Output: Simplified summaries with key obligations, deadlines, and compliance requirements.
    - Purpose: Saves time and removes complexity for legal and compliance teams.

2. **Company-Specific Regulatory Mapping**
    - Tailors compliance guidelines based on company type (e.g., listed/unlisted, SME, mutual fund, stockbroker).
    - Ensures only **relevant regulations** are highlighted, avoiding confusion from non-applicable rules.

3. **Explainability and Source Linking**
    - Every summarized point includes **exact references** (circular number, section, and date).
    - Provides audit-ready evidence, increasing trust among regulators, auditors, and compliance officers.

4. **Continuous Compliance Monitoring**
    - The system will **track new SEBI updates** in real time.
    - Alerts the company about new obligations or changes in existing regulations.
    - Supports **drift detection**: if a regulation changes, it flags what needs to be updated in the company's compliance process.

5. **Search and Query Interface**
    - A chatbot-like interface where users can ask:
        - *"What SEBI rules apply to a listed IT company?*
        - *"What are the disclosure requirements for insider trading?"*
    - The model retrieves relevant SEBI documents and provides clear, sourced answers.

6. **Compliance Checklist Generator**
    - Generates **step-by-step checklists** for a company's compliance obligations.
    - Helps compliance teams track tasks and deadlines.

7. **Risk Reduction & Decision Support**
    - Assists in identifying **potential gaps** in compliance.
    - Provides decision support for corporate governance, reporting, and investor protection measures.

1.3 What We Have to Do (Project Scope)

1. **Collect SEBI Regulations**
    - Gather circulars, guidelines, and reports (from SEBI's official site).
    - Store them in a structured database for retrieval.

2. **Build RAG Pipeline**
    - Use embeddings to store and retrieve SEBI documents.
    - Connect with an LLM (like GPT/LLama) for natural language summarization.

3. **Design Summarization & Mapping Layer**
    - Map SEBI rules to company attributes (listed/unlisted, market cap, industry).
    - Create a logic layer that filters non-relevant regulations.

4. **Develop User Interface**
    - Web application or dashboard named **FinComply**.
    - Features: Upload company details → get compliance summary → interactive chatbot.

5. **Implement Traceability**
    - Every summary or recommendation must include **links to the original SEBI source**.

6. **Enable Alerts & Monitoring**
    - Set up automated checks for new SEBI updates.
    - Notify users of any new obligations via dashboard/alerts.

1.3 Objective

The primary objective of *FinComply* is to simplify and automate compliance with SEBI regulations by leveraging Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs). The system aims to summarize complex SEBI guidelines into simple, actionable insights that are specific to each company's profile, such as whether it is listed or unlisted, its sector, or market capitalization. Another key objective is to ensure company-specific applicability, so that organizations receive only relevant regulatory obligations and avoid unnecessary or non-applicable information. To build trust and audit readiness, *FinComply* will provide transparent outputs with exact references to SEBI circulars, including dates and sections, enabling compliance teams and auditors to verify the source of each recommendation. The system will also focus on real-time updates by continuously ingesting new SEBI circulars and detecting regulatory changes, ensuring that companies remain compliant with evolving rules. Furthermore, the project seeks to enhance decision-making for compliance officers and legal teams by reducing manual effort, minimizing risks of penalties, and saving time. Finally, privacy and security are embedded into the design, ensuring sensitive company data is handled securely while enabling effective compliance automation.

CHAPTER-2

Existing Work / Literature Review

2.1 RAG & Regulatory NLP

Retrieval-Augmented Generation (RAG) combines a retrieval layer (vector search over documents) with an LLM generator to produce grounded, context-aware answers. RAG improves factual grounding and is now the dominant pattern for domain applications (legal, finance, medical). Recent systematic reviews summarize RAG architectures, evaluation challenges, and best practices for domain grounding. arXiv

Regulatory Natural Language Processing (RegNLP) is an emerging subfield focused on processing statutes, circulars, and other regulatory texts. Workshops and community efforts (RegNLP) have crystallized common tasks: obligation extraction, applicability classification, timeline extraction, and provenance-aware summarization. These efforts emphasize domain-specific chunking, legal structure awareness, and evaluation needs. RegNLP+1

2.2 Legal / Financial Document Summarization & RAG

Multiple works adapt RAG specifically for legal texts by using legal-aware chunking, hybrid retrieval + reranking, and constrained generation to reduce hallucinations. These approaches often add a post-hoc provenance step—linking each generated claim to the exact document span—to make outputs more trustworthy for compliance and audit.

Examples include dynamic RAG frameworks and evaluations of LLM robustness in retrieval-based summarization. MDPI+1

Industry legal-tech articles and vendor writeups (e.g., Thomson Reuters, legal tech blogs) point out that domain-specific retrieval (using gold-standard legal corpora) is the key differentiator for trustworthy legal assistants—RAG without domain curation is insufficient. Thomson Reuters Legal

2.3 Temporal / Versioned Legal Corpora and Ontology Approaches

Legal/regulatory corpora are *diachronic* (laws change, circulars are superseded). New architectures model document structure and temporal versions explicitly: ontology-driven or graph RAG architectures treat sections, amendments, and effective dates as first-class entities to avoid recommending obsolete rules. These structure-aware approaches help with “which rule version applies today” queries.

2.4 Privacy & IP Considerations in RAG

RAG systems ingest private or proprietary data (company reports, advisories) and can risk leaking sensitive info via embeddings or generated text. Recent research explores **privacy-preserving RAG**: federated/localized embedding training, homomorphic encryption, and on-prem embedding stores to keep company data secure while still enabling retrieval. These are active research areas with promising prototypes (e.g., federated embedding learning for private RAG).

2.5 SEBI & National Context (Regulator Activity)

SEBI is actively modernizing compliance infrastructure: it has launched unified/digital compliance platforms (e.g., *Samuhik Prativedan Manch*) to simplify reporting for market intermediaries—this shows regulator appetite for tech-enabled compliance solutions. moneylife.in+1

SEBI published a **Consultation Paper (June 20, 2025)** proposing guidelines for responsible AI/ML use in Indian securities markets (model governance, disclosure, testing/monitoring, fairness, data security). This creates both constraints and an opening: any AI compliance assistant must incorporate model governance, explainability, and data security from day one. sebi.gov.in+1

2.6 Benchmarks and Evaluation Work

The community lacks a unified SEBI/India-specific benchmark for regulatory summarization and applicability classification. Research groups propose specialized evaluation suites (factuality + provenance + timeline extraction), but public, domain-specific datasets for SEBI-style tasks remain sparse. This makes cross-paper comparisons hard and motivates dataset release as a contribution. arXiv+1

Novelty

The novelty of *FinComply* lies in its ability to go beyond generic regulatory summarization and deliver **company-specific, audit-ready compliance insights** using advanced LLMs with RAG. Unlike existing research that primarily focuses on retrieving or summarizing legal texts, *FinComply* introduces a **legal applicability taxonomy** that maps SEBI regulations directly to a company's attributes (listed/unlisted, sector, size, intermediary role). This ensures that compliance advice is not only accurate but also highly relevant.

Additionally, the system emphasizes **traceability and provenance**, linking every output to exact SEBI circulars, sections, and dates, thereby making the recommendations verifiable for auditors and regulators—something current tools lack. Another novel contribution is its **real-time lifecycle management**, which continuously ingests SEBI updates, detects regulatory changes, and automatically alerts affected companies. By embedding

privacy-preserving mechanisms in the RAG pipeline, *FinComply* also addresses data security and confidentiality concerns, making it deployable in real corporate environments.

In summary, the uniqueness of *FinComply* lies in combining **applicability mapping, explainability, real-time regulatory tracking, and secure deployment** into one integrated system tailored specifically for SEBI compliance.

3.0 Current Challenges & Gaps (What FinComply Should Target)

3.1 Regulation → Company Mapping (Applicability Taxonomy)

- **Gap**: Most RAG systems return relevant passages but do not systematically decide *whether* a clause applies to a given company profile (listed/unlisted, sector, market cap, intermediary type, cross-border exposure).
- **Why important**: Avoids false positives (irrelevant advice) and false negatives (missed obligations).
- **Opportunity**: Build a hybrid rule + learned applicability classifier tied to company attributes and regulatory ontology.

3.2 Provenance & Audit-Ready Outputs

- Gap: Many summaries lack precise source spans, section/dates, or versioning (which amendment).
- Why important: Regulators/auditors require traceable evidence for any compliance recommendation.
- Opportunity: Enforce source-span linking in generation and present an auditable evidence pane.

3.3 Temporal/version management of regulations

- Gap: Diachronic nature of regulations (amendments, supersessions) is poorly handled by vanilla RAG.
- Why important: Recommending an outdated rule can cause wrong compliance actions.
- Opportunity: Use structure-aware, temporal graph or ontology RAG (e.g., SAT-Graph style) to track versions and effective dates.

3.4 Privacy, IP & Deployability

- Gap: Practical patterns for enterprise privacy (on-prem embeddings, federated retrievers) are still being matured.
- Why important: Firms will not use cloud RAG that risks exposure of sensitive filings or legal strategies.
- Opportunity: Implement privacy-preserving retrieval (local embeddings, federated learning, encryption) and document this as part of the toolchain.

3.5 Evaluation & Benchmarks for SEBI-style Compliance

- Gap: No widely-accepted SEBI or India-specific benchmark measuring applicability precision/recall, timeline extraction, and provenance precision.
- Why important: Without metrics, it's hard to demonstrate legal-grade performance.
- Opportunity: Create and publish a small annotated dataset (applicability + obligations + timelines) as part of the project.

3.6 Model Governance & Regulatory Acceptance

- Gap: Few practical examples showing how an AI compliance assistant meets governance rules (testing, monitoring, human-in-loop). SEBI's consultation makes this requirement explicit.
- Why important: For adoption, FinComply must implement validation, logging, human review workflows, and disclosures about model use.
- Opportunity: Bake governance (testing, model cards, audit logs, human approval gates) into the product by design.

CHAPTER - 3

Proposed Work (Retrieval-Augmented Generation (RAG) based compliance assistant)

The proposed method for *FinComply* involves building a **Retrieval-Augmented Generation (RAG) based compliance assistant** that processes SEBI regulations and delivers **company-specific, explainable compliance insights**. The workflow can be divided into the following stages:

3.1 Data Collection & Preprocessing

- Collect SEBI circulars, guidelines, and regulations from official sources.
- Convert documents into a structured format (text + metadata like date, circular number, section).
- Clean and preprocess text to remove redundancies while preserving legal references.

3.2 Knowledge Base Construction

- Create a vector database using embeddings for efficient semantic retrieval of SEBI rules.
- Tag each regulation with **applicability attributes** (e.g., listed/unlisted, market cap, industry sector, intermediary type).
- Enable taxonomy mapping so that only relevant obligations are retrieved for a given company.

3.3 Retrieval-Augmented Generation (RAG) Pipeline

- When a query or compliance check is initiated, the system retrieves the **most relevant SEBI clauses** from the vector database.
- Use an LLM to generate **summarized, simplified compliance guidance** tailored to the company's profile.
- Ensure each response is accompanied by **traceable citations** (circular number, date, section).

3.4 Company-Specific Customization

- Collect company details (e.g., type, size, listing status).

- Filter retrieved SEBI obligations to ensure only **applicable rules** are presented.
- Highlight non-compliance risks or upcoming regulatory deadlines.

3.5 Real-Time Updates & Drift Detection

- Continuously monitor SEBI's website for new circulars or amendments.
- Automatically re-ingest and re-index updated documents.
- Alert companies about **changes that directly impact them**.

3.6 Explainability & User Interface

- Provide a **dashboard** for compliance teams with summarized obligations, applicability tags, and source references.
- Enable **downloadable compliance reports** for audit purposes.
- Offer explainable AI outputs so that compliance officers can trust and verify the recommendations.

3.7 Privacy & Security Layer

- Implement on-premise or private cloud deployment to prevent leakage of company-sensitive data.
- Use **redaction/obfuscation techniques** for sensitive fields during query processing.

Expected Outcome

The proposed solution will be developed iteratively following the Agile methodology, ensuring that the system evolves from a lightweight prototype into a fully integrated compliance automation platform. The anticipated outcomes across the various phases are outlined below:

Phase I – Prototype Development

In the initial phase, a proof-of-concept will be built to demonstrate the feasibility of the system. A basic retrieval-augmented generation (RAG) pipeline will be implemented, trained on SEBI circulars, guidelines, and compliance frameworks. A simple web interface will be provided, allowing developers to paste or upload code and documentation snippets. The system will generate text-based compliance guidance, flagging potential non-adherence to regulatory requirements. At this stage, the solution will serve as an assistive tool for developers and will not be connected to CI/CD pipelines.

Phase II – Developer Assistance Tool

The second phase will focus on enhancing usability for developers. Document upload and parsing capabilities will be introduced, enabling contextual compliance analysis. The system will also provide interactive responses through a chatbot interface, offering clarifications such as “What are SEBI guidelines on data privacy for fintech applications?”. A feedback mechanism will be integrated to allow developers to evaluate suggestions, thereby improving the model’s accuracy. The outcome of this phase will be a developer-focused compliance advisor that can assist in creating policy-compliant code and documentation. However, integration into developer workflows will remain manual.

Phase III – Pre-Integration Enhancements

The third phase will expand the solution’s scope to organizational use cases. A compliance dashboard will be developed to provide visual indicators of compliance status using a traffic-light model (green/yellow/red). Automated monitoring will be added to update the knowledge base as SEBI releases new circulars and regulatory amendments. Additionally, the system will be capable of generating company-specific compliance reports, identifying adherence levels, and highlighting regulatory gaps. At this stage, the platform will serve as an organizational compliance assistant, semi-automated and primarily intended for review and internal evaluation.

Phase IV – CI/CD Integration and Automation

The final phase will involve embedding the compliance assistant directly into CI/CD pipelines such as GitHub Actions, GitLab CI, or Jenkins. The system will perform automated compliance checks during code commits and pull requests, generating detailed reports that highlight violations and suggest corrective actions. Furthermore, audit trail generation will be implemented, enabling the export of compliance reports in standardized formats (e.g., PDF, Excel) for review by compliance officers. The final deliverable will be a fully automated, enterprise-ready compliance solution integrated into the software development lifecycle.

Final Outcome

By following this phased approach, the project will evolve from a prototype assistive tool into a robust, AI-driven compliance automation platform. The end system will enable fintech developers and organizations to ensure continuous adherence to SEBI regulations, thereby reducing compliance risks, minimizing manual oversight, and fostering regulatory trust in fintech innovations.

REFERENCES

1. Umar, M., Siddique, R., Liew, C., Ali, M., Habib, S. 2025. Enhancing Regulatory Compliance Through Automated Retrieval, Reranking, and Answer Generation. *Proceedings of the Workshop on Natural Legal Language Processing (RegNLP 2025)*. ACL Anthology.

2. Purbey, S., Garg, D., Malhotra, S., Luthra, P., Choudhary, A. 2024. 1-800-SHARED-TASKS at RegNLP: Lexical Reranking of Semantic Retrieval (LeSeR) for Regulatory Question Answering. *arXiv preprint arXiv:2412.06009*.

3. Parikh, A., Mathur, P., Mehta, H., Mittal, A., Majumder, P. 2021. LawSum: A Weakly Supervised Approach for Indian Legal Document Summarization. *arXiv preprint arXiv:2110.01188*.

4. Shreyasi, B., Bandyopadhyay, S., Chakrabarti, S. 2025. Extractive Summarization of Indian Legal Judgments Using Unsupervised Role Labelling. *Artificial Intelligence and Law*. Springer.

5. Sun, Z., Yu, D., Hou, Y., Lin, C., Liu, X. 2025. A Two-Stage LLM System for Enhanced Regulatory Information Retrieval and Answer Generation. *Proceedings of the Workshop on Natural Legal Language Processing (RegNLP 2025)*. ACL Anthology.