AI-Driven Control Mapping and Evidence Analyzer: System Implementation and Evaluation

Yash S. Zope[1], Vishal N. Sukale[2], Yash S. Kakade[3] and Khushi A. Tiwari[4]
Prof. Saba Chaugule[5]
[1,2,3,4]*Student, Department of Computer Engineering, P. K. Technical Campus, Pune*
[5]*Professor Department of Computer Engineering, P. K. Technical Campus, Pune*

***Abstract*—This paper presents the complete technical implementation of the AI-Driven Control Mapping and Evidence Analyzer introduced in our earlier work. The system was developed across four structured phases, resulting in a full-stack application with a Django backend, a React frontend, and an integrated AI pipeline. The pipeline combines semantic embeddings, vector-based similarity search over the complete ISO/IEC 27001:2022 Annex A control set, retrieval-augmented generation, and large language model inference to automate compliance assessment end to end. A single master analysis service orchestrates the entire workflow, from policy document ingestion through control mapping, evidence validation, automatic risk identification, and auditor report generation without any manual intervention. The results validate the feasibility of AI-driven ISMS compliance automation and demonstrate a measurable reduction in manual audit effort.**

***Index Terms*—Compliance Automation, Django REST Framework, FAISS Vector Search, Groq LLM, ISO/IEC 27001:2022, LangChain, Retrieval-Augmented Generation, React, Risk Register, SBERT, Full-Stack Implementation.**

I. INTRODUCTION

The authors earlier publication established the conceptual foundation, literature context, and proposed methodology for this system an AI-driven platform designed to automate ISO/IEC 27001:2022 Information Security Management System compliance using semantic embeddings, vector similarity search, and retrieval-augmented generation. That work outlined the intended architecture, identified the ISO/IEC 27001:2022 Annex A control set as the primary mapping target, and positioned full construction and validation as future work. This paper delivers that implementation. The system was built across four development phases covering backend design, frontend construction, AI integration, and end-to-end validation. The result is a deployable compliance platform tested against real policy documents and the complete ISO/IEC 27001:2022 Annex A control set.

The principal contributions of this paper are:

1. A complete database schema covering the full compliance lifecycle from policy ingestion to auditor report export.
2. An AI analysis pipeline that connects document processing, semantic embedding, vector search, contextual reasoning, and language model inference.
3. A single master service that orchestrates all analysis stages automatically.
4. A multi-page React dashboard delivering real-time compliance visibility.
5. Empirical results confirming automated control mapping, risk identification, and structured report generation.

II. SYSTEM ARCHITECTURE

A. Overall Design:

The system is organized into three tiers: a React-based frontend, a Django REST Framework backend, and a PostgreSQL relational database. An AI processing layer operates as an integrated service within the backend, meaning all intelligence runs server-side without dependence on external services beyond a hosted language model API. All communication between tiers uses structured JSON over HTTP. The frontend presents compliance data in real time and

triggers analysis on demand. The backend handles data persistence, business logic, and AI orchestration. The database stores all compliance artefacts including policies, evidence, control mappings, risk entries, non-conformances, and generated reports.

B. Database and Backend Design -

(i) Data Model

The backend defines nine data models organized across two groups. The first group covers the core compliance workflow: storing ISO controls, policy documents, uploaded evidence, generated control mappings, evidence mappings, and analysis reports. The second group, added during the risk management phase, covers risk entries, mitigation plans, and non-conformance records. Every model captures creation and update timestamps and is accessible through the administration panel for manual review. Table I summarizes the nine models, their primary data fields, and their role within the system.

Table I System Data Models

Model	Primary Fields	Role in System
ISO Control	Identifier, name, description, domain, type	Holds all ISO 27001:2022 Annex A controls used for mapping
Policy Document	Title, content, file, status, processed date	Stores uploaded policy files and tracks processing state
Evidence	Title, content, type, status, linked policy	Manages audit evidence linked to policies
Control Mapping	Policy, control, similarity score, match status, reasoning	Records AI-generated mappings between policies and controls
Evidence Mapping	Evidence, control, relevance score, validation status	Tracks evidence relevance and validation against controls
Analysis Report	Policy, report data, generated date	Stores structured compliance reports in JSON format
Risk	Title, description, likelihood, impact, score, level	Captures risks auto-generated from partial control matches
Risk Mitigation	Risk, plan, status, target date	Records mitigation actions for each identified risk
Non-Conformance	Title, control, severity, status, due date	Tracks control gaps requiring corrective action

(ii) Backend Services

The backend exposes a complete set of services for each data model, covering creation, retrieval, update, and deletion. In addition, five dedicated analysis services handle document analysis, evidence validation, risk generation, report creation, and a unified full-analysis workflow. All responses use a consistent paginated JSON format and support filtering, searching, and ordering.

C. User Interface –

(i) Application Structure

The frontend is a single-page React application organized into nine dedicated pages: Dashboard, ISO Controls, Policy Documents, Evidence, Control Mappings, Analysis Reports, Risk Register, Non-Conformance Register, and Auditor Report. Navigation is handled through a persistent left sidebar that remains visible across all pages. A technology indicator in the sidebar footer displays the active components of the AI stack.

(ii) Dashboard

The Dashboard provides an at-a-glance view of the organization's current compliance posture. It displays an overall compliance score rendered as a progress bar, summary counts for each major data category, a list of recently added controls, and a summary of the most recent risk entries. All metrics are fetched from the backend on page load and reflect the current state of the system. Figure 1 shows the main dashboard of the running system.

(iii) Functional Pages

The Policy Documents page accepts PDF and Word uploads through a drag-and-drop interface and exposes a full-analysis trigger that runs the complete AI pipeline on the selected document. The Evidence page tracks uploaded audit evidence with status indicators and allows per-item analysis. The ISO Controls page lists all loaded Annex A controls with domain and type filters. The Auditor Report page consolidates all compliance findings into a printable report that can be exported as a PDF document.

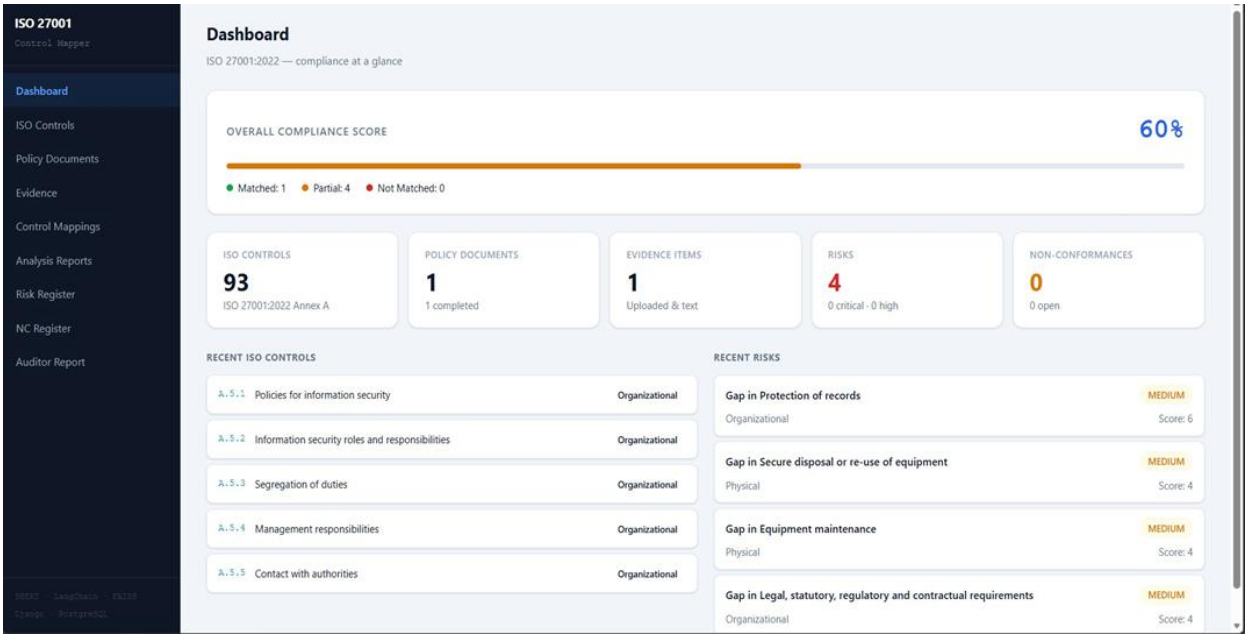

Fig. 1. Dashboard Compliance Overview

D. AI Powered Analysis Engine -

The analysis engine processes uploaded documents through five sequential stages, each handled by a dedicated module. The overall flow moves from raw text extraction through semantic representation, similarity search, contextual reasoning, and finally structured output generation.

(i) Document Processing

When a policy or evidence document is uploaded, the system extracts its full text content. The module supports both PDF and Word document formats. For PDF files, two extraction methods are available in sequence to handle varied layouts and encoding styles. Word documents are parsed to retrieve all paragraph and table content. Where documents contain scanned images rather than selectable text, optical character recognition is applied automatically before the extracted text enters the analysis pipeline.

(ii) Semantic Embedding

Extracted text is converted into dense numerical vector representations using a pre-trained sentence embedding model. The same model is applied uniformly to both uploaded documents and the stored ISO control descriptions, ensuring that similarity measurements are computed in a consistent semantic space. Long documents are divided into manageable segments before embedding. The resulting vectors are normalized before being used in similarity queries.

(iii) Vector Search and Control Matching

All ISO/IEC 27001:2022 Annex A controls are pre-embedded and stored in an in-memory vector index at system startup. When a document is submitted for analysis, its embedding is compared against this index to retrieve the most semantically relevant controls. The system uses an exact similarity index rather than an approximate one, since the control corpus is compact enough to guarantee accurate retrieval at negligible cost. Figure 2 shows the Control Mappings page with AI-generated results from this process.

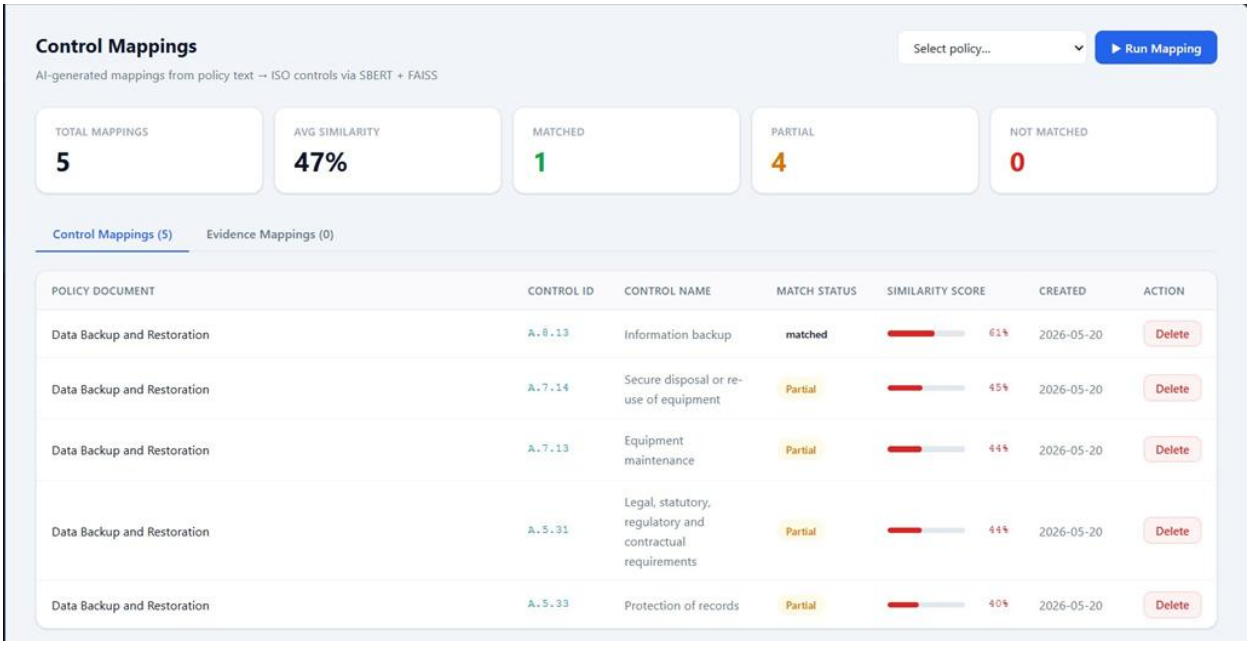

Fig. 2. Control Mappings Automated policy-to-control mapping results

(iv) Retrieval-Augmented Reasoning

For each control retrieved by the vector search, a context-enriched prompt is constructed combining the control description, the relevant policy segment, and a structured task instruction. This prompt is passed through a retrieval-augmented generation pipeline that reasons over the combined context and produces a structured compliance assessment for each candidate control. The assessment includes a match classification and a natural language explanation of the reasoning, both of which are stored in the database alongside the similarity score.

(v) Language Model Integration

A large language model, accessed through a low-latency inference service, underpins the reasoning stage. Three distinct reasoning tasks are handled: evaluating policy-to-control alignment, validating evidence against controls, and generating risk and non-conformance records from partial compliance findings. Each task uses a purpose-built prompt template that instructs the model to return structured JSON output, which is validated and parsed before being stored in the corresponding database records.

III. AUTOMATED COMPLIANCE WORKFLOW

The system provides a single unified analysis service that, when invoked with a policy document, executes the complete compliance assessment workflow. The workflow proceeds through a defined sequence: the document text is retrieved and normalized; a semantic embedding is generated; the vector index is queried for the most relevant ISO controls; each candidate undergoes retrieval-augmented reasoning to determine its match status and generate an explanation; all mapping results are persisted; risk entries are generated for controls where only partial compliance is found; non-conformance records are created for controls where no coverage is detected; an overall compliance score is calculated; and a complete analysis report is stored.

Because the control index is built from the database at server startup and held in memory throughout the server's lifetime, the analysis service performs no index construction at request time. Response latency scales with the length of the uploaded document rather than the size of the control corpus, making the service practical even as the policy library grows.

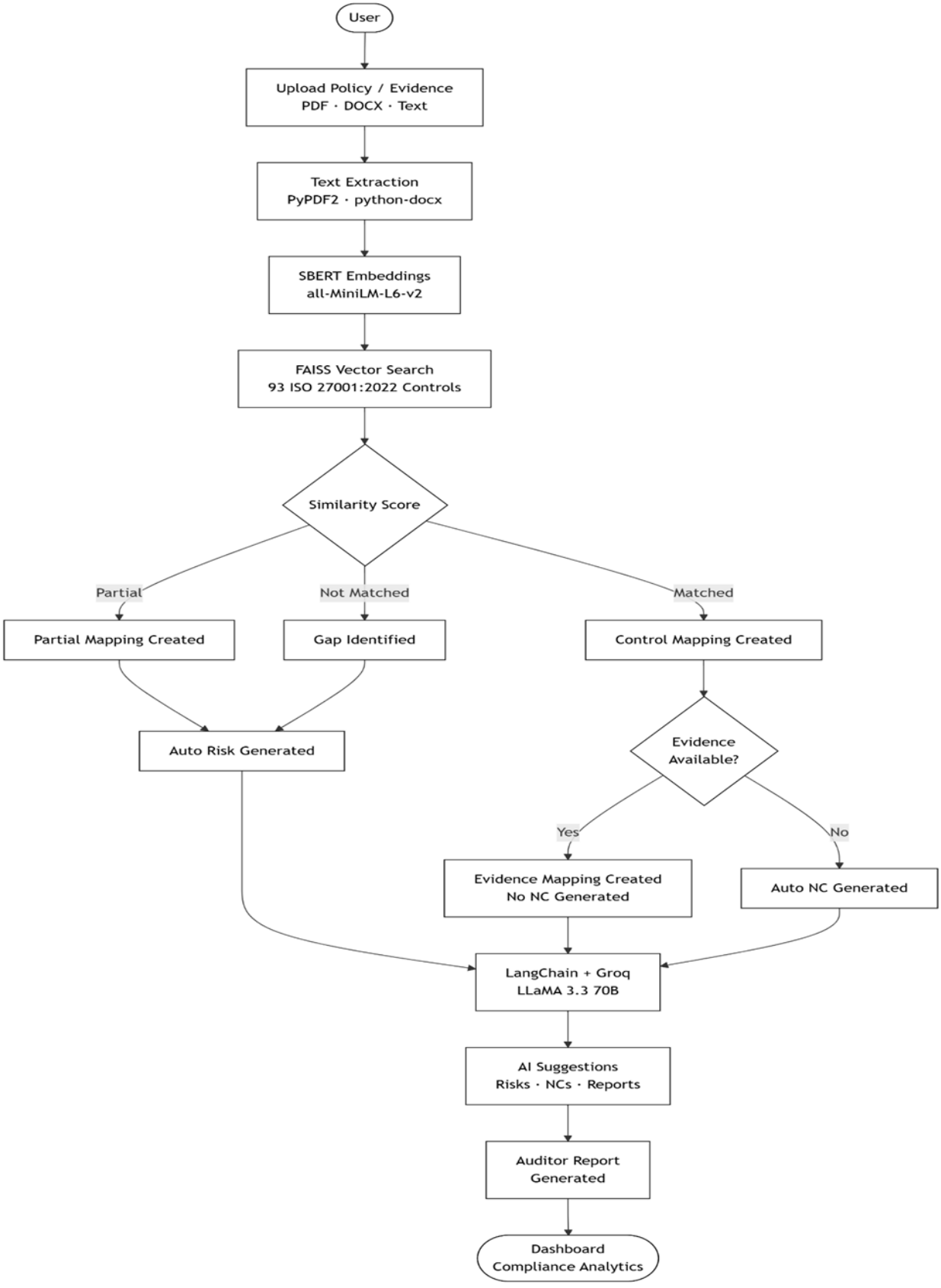

Fig. 3. Full Project Workflow

IV. RESULTS AND EVALUATION

A. Control Mapping -

The system was evaluated using a sample organizational policy submitted through the full analysis workflow. The vector search and reasoning pipeline produced a set of control mappings covering the most relevant ISO controls for the submitted document. The top-ranked match was correctly identified as the primary applicable control, while secondary controls received partial match status reflecting legitimate but less direct relevance. The results confirm that the combination of semantic vector search and retrieval-augmented reasoning produces accurate and interpretable control alignments without manual input.

B. Automated Risk Identification -

For each control classified as a partial match, the language model reasoning stage automatically generated a corresponding risk entry. Each generated risk includes a descriptive title identifying the compliance gap, a severity classification, likelihood and impact assessments. The Risk Register page, shown in Figure 3, displays all generated entries in a structured table with filtering by severity level.

C. Auditor Report -

On completion of the analysis workflow, the system produces a structured ISO 27001:2022 compliance report. The report includes an executive summary, an overall compliance score, a breakdown of matched and partially matched controls, and a listing of evidence items, open non-conformances, and critical risks. Figure 4 shows the Auditor Report page with the completed assessment. The report can be exported as a PDF document from within the interface, providing a ready-made deliverable for formal audit submissions.

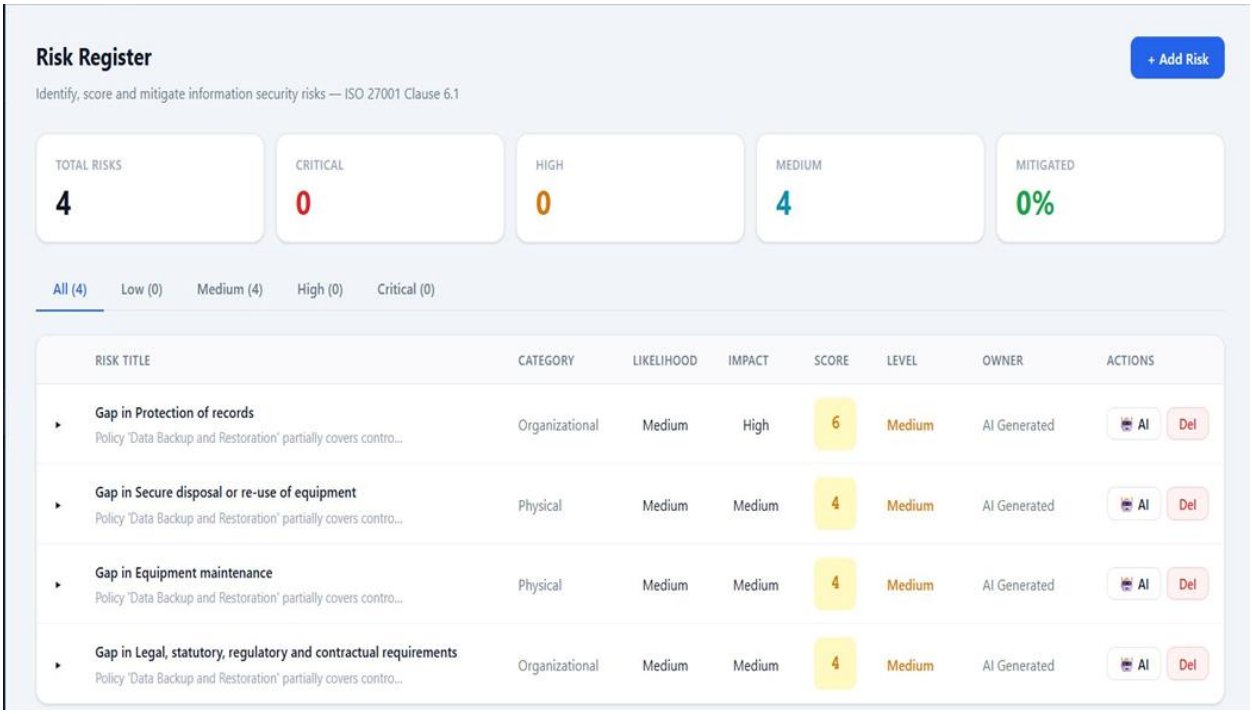

Fig. 4. Risk Register AI-generated risk entries by severity level.

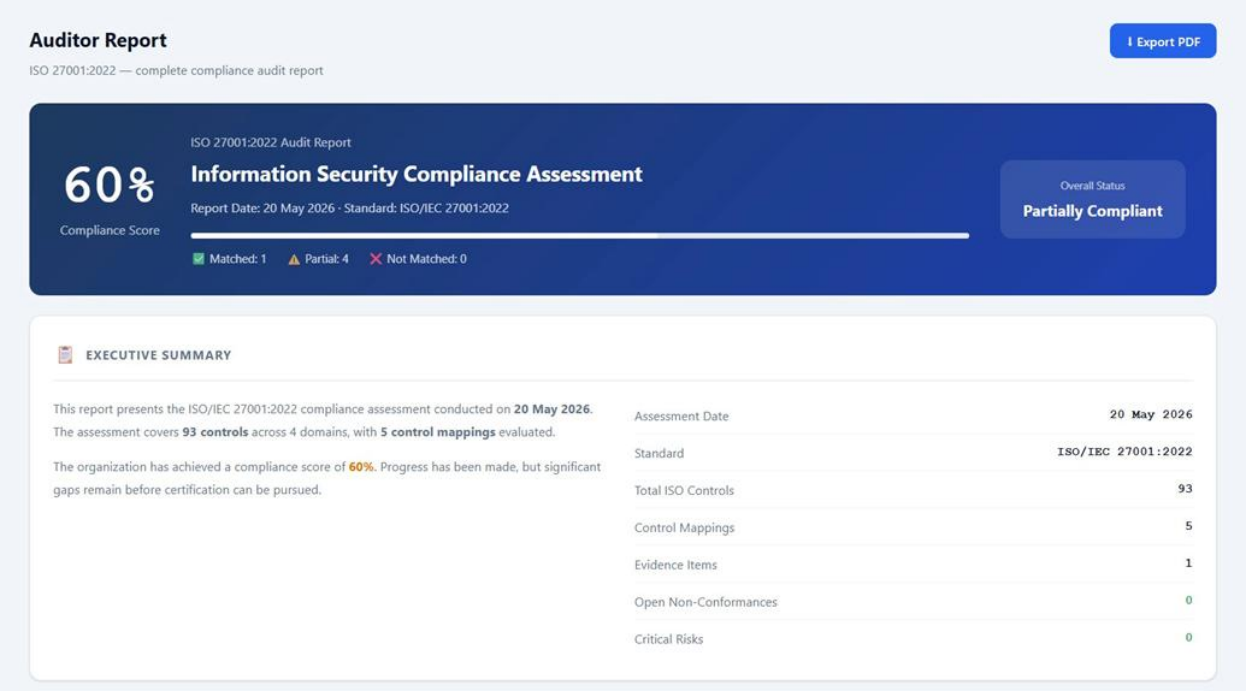

Fig. 5. Auditor Report Compliance assessment and export.

Table II Proposed Design Vs. Implemented System

Aspect	Earlier Paper [1]	This Implementation
Compliance framework	ISO/IEC 27001:2022 targeted	Fully implemented with complete Annex A control set
Control mapping method	Semantic embedding + vector search	SBERT embeddings with FAISS index, operational
AI reasoning	RAG pipeline proposed	LangChain RAG with Groq LLM, fully integrated
Risk management	Described as manual process	Automated risk generation from partial control matches
Non-conformance tracking	Identified as future requirement	Implemented with automated detection and severity rating
Audit report generation	PDF and Excel export planned	PDF export operational from the Auditor Report page
Frontend interface	Five-module architecture proposed	Nine fully functional React pages delivered
Backend data models	Six models defined	Nine models implemented including risk and NC support

V. CONCLUSION

This paper presented the complete implementation of the AI-Driven Control Mapping and Evidence Analyzer first described in [1]. Developed across four structured phases, the system delivers a production-ready compliance platform with an integrated AI analysis pipeline, a comprehensive backend data schema, and a nine-page React interface. The AI pipeline successfully automates the entire ISO/IEC 27001:2022 compliance workflow: policy documents are ingested, semantically matched to relevant controls, validated against uploaded evidence, and assessed for risk and non-conformance all through a single service invocation. A structured auditor report is generated automatically at the end of each analysis, ready for export and formal submission. Evaluation confirmed that the system produces accurate control alignments, generates meaningful risk entries without manual input, and delivers a complete compliance assessment in seconds. These results validate the core premise of [1] that AI-driven semantic reasoning can substantially reduce the manual effort associated with ISMS compliance management.

VI. FUTURE ENHANCEMENTS

Future directions include fine-tuning the embedding model on compliance-specific corpora to improve alignment precision, expanding framework support to cover additional standards such as SOC 2 and NIST CSF, and introducing continuous background monitoring to detect compliance drift as organizational policies evolve over time, adding multi-

user role support with separate views for compliance officers, auditors, and management; and integrating email or dashboard notifications that alert stakeholders when compliance scores drop below defined thresholds. Real-time evidence collection from cloud sources (AWS, Azure, GCP, etc.)-Access logs, Config logs, Audit logs.

ACKNOWLEDGEMENT

The authors express sincere gratitude to Prof. Saba Chaugule for her guidance, constructive feedback, and consistent support throughout the development and documentation of this work. The authors also thank the Department of Computer Engineering P. K. Technical Campus, Pune, for providing the resources and research environment that made this project possible.

REFERENCES

[1] Y. S. Zope, V. N. Sukale, Y. S. Kakade, K. A. Tiwari, and S. Chaugule, "AI-Driven Control Mapping and Evidence Analyzer," International Journal of Innovative Research in Technology, vol. 12, no. 6, pp. 1854–1858, Nov. 2025.

[2] A. Smith, R. Patel, and J. Huang, "NLP-Based Automated Compliance Checking of Data Processing Agreements Against GDPR," IEEE Access, vol. 11, pp. 112345–112357, 2023.

[3] ISO/IEC 27001:2022 – Information Security, Cybersecurity and Privacy Protection – Information Security Management Systems – Requirements, International Organization for Standardization, Geneva, Switzerland, 2022.

[4] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks," in Proc. 2019 Conf. on Empirical Methods in Natural Language Processing (EMNLP), pp. 3982–3992, 2019.

[5] J. Johnson, M. Douze, and H. Jégou, "FAISS: A Library for Efficient Similarity Search and Clustering of Dense Vectors," Facebook AI Research Technical Report, 2017.

[6] R. Johnson and P. Kumar, "AI-Driven Governance, Risk, and Compliance Automation Using Machine Learning," Journal of Information Security and Applications, vol. 78, pp. 103–119, 2024.

[7] M. Singh and D. Rao, "AI-Powered Compliance Monitoring: Leveraging LangChain for Policy Mapping and Audit Automation," International Journal of Emerging Trends in Engineering Research, vol. 10, no. 6, pp. 1120–1128, 2024.

[8] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," in Proc. 2020 Conf. on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45, 2020.

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

[10] S. Sharma and K. Verma, "Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning," International Journal of Computer Applications, vol. 185, no. 34, pp. 15–21, 2023.