INTEGRATION OF AI TECHNOLOGIES IN BUSINESSES AND ITS IMPLICATION TO CORPORATE RISK MANAGEMENT MODEL - A CASE STUDY ON THE IMPLEMENTATION OF RAG APPLICATION AND RISK ANALYSIS IN THE CONTEXT OF 3 LINES OF DEFENSE

[1]JIANG HE, [2]JING LI

[1]Ph.D., York University, Toronto, Canada
[2]Ph.D., Independent Researcher, Toronto, Canada
E-mail: hejiang@yorku.ca

Abstract - Organizations are rapidly adopting retrieval-augmented generation (RAG) to improve access to internal knowledge and enhance operational efficiency, yet most lack mature frameworks to manage the unique risks introduced by AI systems. This study examines the integration of RAG into enterprise environments through a detailed case analysis of a mid-size Canadian credit union, whose existing risk governance follows the Three Lines of Defense (3LOD) model but does not explicitly recognize AI as a risk category. Using the MIT AI Risk Repository as a reference taxonomy, we identify nine risk scenarios across the RAG pipeline—from ingestion and embeddings to retrieval, generation, and drift monitoring—and evaluate how each can be addressed within the 3LOD structure. The case study demonstrates that traditional controls designed for non-AI systems are insufficient for managing risks associated with AI solutions (e.g., hallucination, traceability failures, embedding manipulation, and AI-system performance drift, etc.). We propose both short-term mitigation strategies and longer-term governance enhancements to elevate organizational maturity in Responsible AI. The findings highlight that while AI risks can be integrated into existing enterprise risk processes, organizations must conduct comprehensive gap analyses and update governance mechanisms to ensure AI use cases meet business objectives while complying with internal policies and external regulations.

I. INTRODUCTION

As organizations increasingly rely on large language models, retrieval-augmented generation (RAG) has gained traction in enterprise settings because it grounds model outputs in validated, internally governed knowledge bases.

At its core, retrieval-augmented generation (RAG) is an AI system pattern that connects a large language model (LLM) with a retrieval layer that fetches relevant, up-to-date context from a trusted knowledge base at query time. Instead of relying solely on what the model learned during pre-training, RAG "grounds" responses in specific source materials (documents, databases, APIs), improving factual accuracy, timeliness, and verifiability.

Leveraging a case study, this paper discusses the risks associated with the implementation of RAG in an enterprise environment.

Particularly, we discuss its risk implications in the context of 3 lines of defense, a widely adopted risk management model, and propose mitigating solutions to address such risks.

To ensure our findings and recommendations can be applied to a wider audience and risk practitioners, this paper reviews the risk findings and proposed solutions leveraging the MIT AI Risk Repository (Slattery et al., 2026), a leading AI risk taxonomy which has been recognized by the industry.

This paper is intended to address a growing gap between fast AI innovation and lagging risk governance framework by mapping latest AI risk management principles to real-world AI use cases. And our risk analysis performed in the context of 3-line of defense, an industry-wide risk governance model, provides important learnings for organizations to upgrade their risk management program in response to the new reality brought by AI.

1.1. Retrieval-augmented Generation Workflow

Compared with standalone LLM use, the RAG system added three additional key layers between the user input (i.e., prompts) and system output:

Interpret Query – Interpret user requirements in the business context

Retrieval – System retrieves data from trusted knowledge base

Prompt Augmentation – System produces an augmented prompt to the LLM with enhanced context from the retrieved data

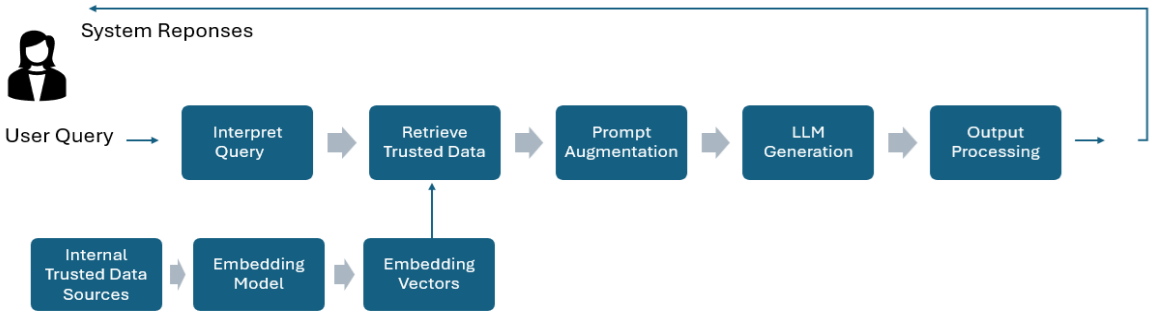


Figure 1. High-level Workflow of RAG

1.2. Benefits of RAG for Enterprise Users

Enterprises often store policies, procedures, and standards across heterogeneous repositories (shared drives, SharePoint, wikis, ticketing systems) owned by different teams. RAG enables a unified, permission-aware interface that surfaces the "right paragraph" with traceable citations rather than brittle keyword search or manual spelunking. When implemented appropriately, it provides an important tool for users to access the up-to-date and accurate business information leveraging the flexible user interface enabled by LLM.

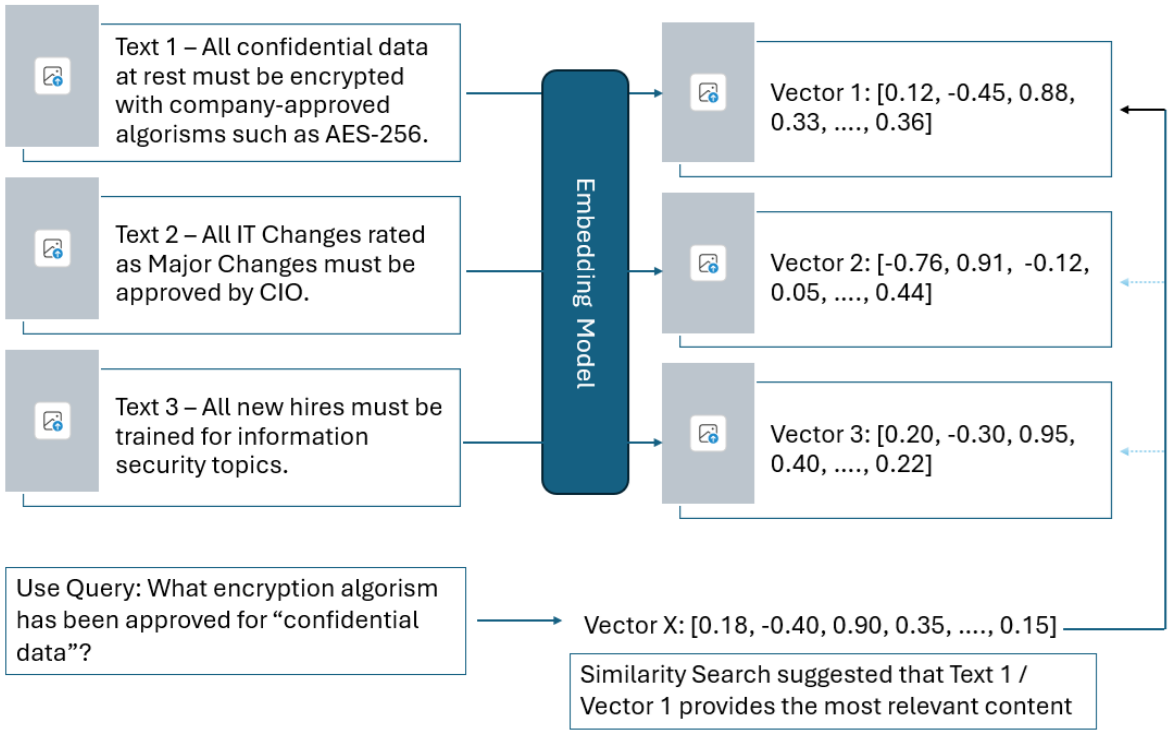


Figure 2. A Simplified Example of Similarity Search Leveraging Vector Databases

Second, by constraining answers to retrieved, governed sources—and showing citations—RAG builds trust with risk, audit, and legal stakeholders. As a result, the risk of hallucinations can be reduced significantly.

Third, the implementation of RAG would require businesses to examine their data governance practices first, allowing many of the data quality and security issues to be identified and remediated at sources. RAG pipelines provide great opportunities to normalize content at ingest, apply quality and security checks, helping organizations meet the increasingly stringent data protection and privacy regulations.

1.3. Necessity of the study supported by literature review

Although RAG solutions provide many benefits to enterprise users, their implementation may bring significant challenges in risk management for those organizations which have not established a dedicated AI Risk Management framework.

The academic research about AI risk assessment in business context is an emerging but rapidly growing research area.

Bibliometrics provide a quantitative method to analyze scientific dynamics and trends in a specific research area.Based on a bibliometric analysis, a research team (Muria-Tarazon, et al., 2025) scanned the in-scope academic sources between 2018 and 2024 covering more than 22,000 peer-reviewed journals, and confirmed that AI risk assessment in business is a very novel research subject, in which explainable AI and cyber security being the most actively researchedareas. Due to its novelty, there is little existing in the body of literature discussing the implication of AI risks to the traditional risk governance model.

The low volume of rigorous research output in this area can be partially attributed to the speed of AI development. The recent acceleration in AI advancement, which followed the appearance of LLM-based GenAI, has been astonishing for both the academic community as well as risk management practitioners.

A 2025 industry survey supported by 233 security, compliance and data leaders (BigID, 2025), concluded that 64.4% of organizations lack full confidence in securing AI-driven data, and 69.5% of organizations cite AI-powered data leaks as their top security concerns in 2025; meanwhile 47.2% suggested that they have no AI-specific security controls in place. The results from the industry survey are in line with our literature study outcome based on the academic sources.

II. OVERVIEW OF CASE STUDY AND RESEARCH METHODOLOGY

2.1. Background of the Organization and Regulatory Landscape

Our study is based on a case analysis with a mid-size Credit Union operating in Ontario, Canada, which offers a variety of financial services products in the province. Due to confidentiality reasons, we are not disclosing its real name and “Firm X” is used to represent the business throughout this research article.

As a provincially regulated Credit Union, Firm X is not mandated to meet the security, operational resiliency, and model risk requirements defined by the country’s financial regulator (i.e., Office of the Superintendent of Financial Institutions), but its data handling practices are directly subject to the country’s data privacy regulations including but not limited to the Personal Information Protection and Electronic Documents Act (PIPEDA) (Government of Canada, 2000) and other provincial-level data privacy regulations.

At the time of conducting this research, there were no AI-specific regulations or legislations directly applicable to the business. As the business has no operations in EU, the EU AI Act (European Commission, 2025) is not applicable.

2.2. Overview of Research Methodology

This research adopts a qualitative case study methodology to examine the integration of retrieval-augmented generation (RAG) technology into an enterprise environment and its implications for risk governance under the Three Lines of Defense (3LOD) framework. Case study approach is particularly suitable for investigating contemporary phenomena within their real-world organizational context, especially when the boundaries between the phenomenon and its context are not clearly evident (Yin, 2018).

The implementation of artificial intelligence (AI) solutions in organizations represents a complex socio-technical phenomenon involving technology, governance structures, risk management processes, organizational culture, and regulatory considerations. Because many of these factors interact dynamically and cannot be isolated in a controlled experimental setting, a case study approach provides an appropriate methodology for generating context-rich insights and understanding how AI-related risks emerge and are managed in practice.

This study follows a single-case, exploratory design. Raw data for the case analysis was collected in Q4 2025, based on interviews with key stakeholders from the Firm X. The RAG implementation itself was conducted throughout Q1-Q3 of 2025.

The analysis focused on identifying risk scenarios across the end-to-end RAG lifecycle and examining how these risks were assessed, assigned, monitored, and mitigated within the existing risk governance structure.

Therisk mapping to the MIT AI Risk Repository (Slattery et al., 2026) enabled the researchers to compare organization-specific observations against a broader industry-recognized framework and to identify potential governance gaps that may be applicable beyond the focal organization.

In organizations where “AI risk” is not yet explicit in their risk taxonomy, RAG provides a concrete entry point: it touches well-understood categories (e.g., information security, data privacy, model risk, operational risk, legal/compliance, etc.) and yields measurable risk metrics tied to real workflows (e.g., percent of answers with valid citations; percentage of

queries blocked by permission filters; rate of red-team prompt injection detections, etc.).

The objective of this study is not statistical generalization but analytical generalization. Consistent with established case study methodology, our findings are intended to contribute to theory and practice by illustrating how existing risk management frameworks can be adapted to address emerging AI risks.

2.3. AI Use Case being implemented

The Use Case being implemented is a general ChatBot facilitating internal users to achieve the following objective: 1) finding and locating the up-to-date internal policies, procedures, and product manuals, etc. 2) translating documents from English to French, or vice versa, 3) helping end users summarize meeting notes and/or reformat documents.

As the firm's very first major AI use case, the selection and scope of the use case has considered the following key facts which have been validated by internal stakeholders, 1) the in-scope internal artifacts are scattered in a number of fragmented IT systems including some legacy applications, 2) there was a strong demand from employees, supported by the management's goal for enhancing operational efficiency, 3) the firm has embarked its journey for digital transformation and a centralized data lake on AWS is in place capturing a number of mission-critical databases and workloads, 4) similar to other industry peers of similar size, the firm has very limited in-house AI expertise but is willing to engage external resources to support the implementation.

2.4. Risk Governance Program in Place

As a trusted financial services provider, Firm X has established its risk management program following the widely accepted Three Lines of Defense (3LOD) Model (i.e., Business Units / Front-line Employees as the 1st line risk owner; Risk Management, Legal & Compliance as the 2nd line function tasked with oversight responsibilities; Internal Audit as the 3rd line). The diagram below (Figure 3) demonstrates the typical structure of 3 Lines of Defense which is commonly observed in the financial industry. It is noteworthy that different organizations may tailor the governance model to fit with their own organization structures.

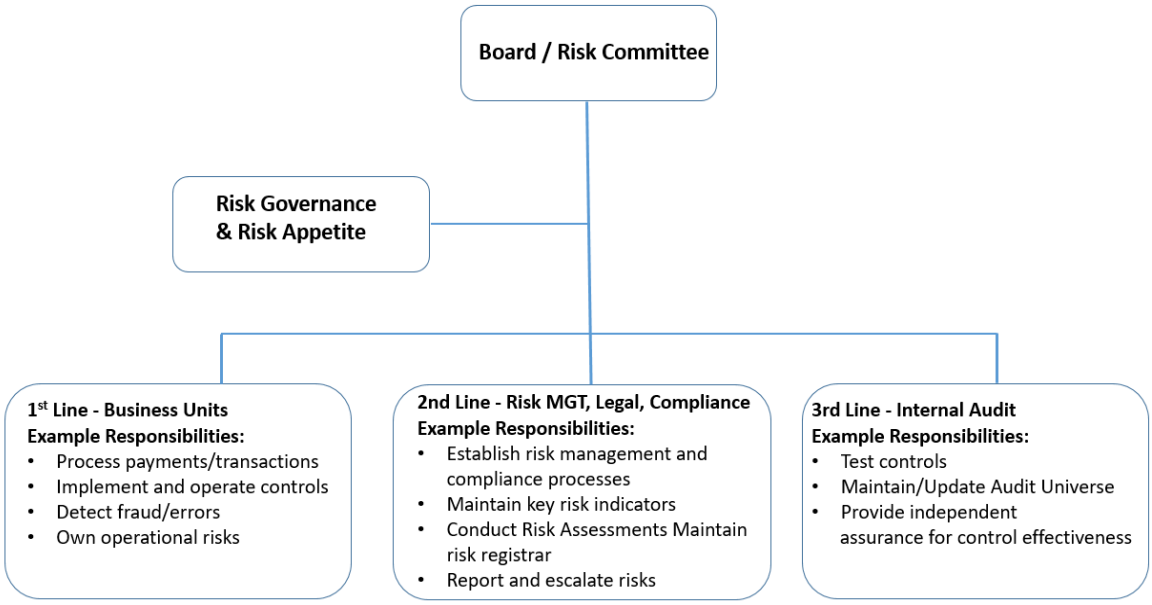


Figure 3. Risk Governance Structure based on 3 Lines of Defense

Although the 3LOD model has been established and functioning well, Firm X's risk taxonomy has not listed AI as a separate risk category, and the most relevant risk categories are 1) operational risk, in which cyber security, data management, information technology, and model risk are listed as sub-categories, and 2) legal & compliance risk, in which privacy risk is listed as a sub-category.

The Operational Risk Management Committee, consisting of senior management and select business unit representatives, performs key oversight function for issues related to operational risks but AI risk has not been explicitly codified into the mandate of the committee.

This type of risk taxonomy setup is common for financial institutions operating in the Canadian market, and the above-mentioned non-financial risk categories & sub-categories have been well aligned with the Basel III Framework (BCBS, 2011) and COSO Enterprise Risk Management Framework (IIA, 2020).

In terms of the firm's maturity level in the context of RAI (Responsible AI), the profile of Firm X matches the "Emerging" category defined by the Microsoft

Responsible AI Maturity Model Microsoft (Vorvoreanu et al., 2023). A key characteristic is that the management has not officially embedded the RAI into their decision-making process though the internal stakeholders have reached a consensus that AI needs to be managed carefully with a consideration of industry best practices while satisfying relevant compliance requirements.

III. THREATS AND RISKS FOR RAG USE CASES

3.1. Development of Risk Scenarios

Given the background information explained in Section 2.2 and 2.3, Firm X managed to revise their existing governance processes to evaluate the risks associated with the AI use case. In this section, we explain in detail how the AI-specific risks along with generic technology & cyber risks were addressed using the revised governance model under 3 Lines of Defense.

In a typical enterprise deployment, the loop of RAG sits inside an end-to-end pipeline: ingestion → indexing → retrieval → re-ranking → context assembly → generation → evaluation → feedback.

In the following, we break down the operational risks associated with the ChatBot implementation by the loop, and discuss their implications for the existing risk governance model.

To increase the generalizability of our case study, the identified operational risks are mapped onto the MIT AI Risk Repository (Slattery et al., 2026), an industry-leading risk taxonomy specifically developed for AI solutions.

	Key Tasks to be Achieved	**Risks Scenarios Identified through Preliminary Analysis**	**MIT AI Risk Domain**
Data ingestion and normalization	- Parse varied formats (PDFs, office docs, wikis, emails, tickets, logs, tables). - Preserve structure (headings, tables, figure captions) and capture provenance (source, owner, timestamps, change history). - Enforce security tagging (sensitivity, access control lists), jurisdiction and lifecycle metadata (effective/expiry dates).	Data quality issues at the sources prevent the model from producing reliable output; and specifically - Metadata (e.g., ownership, security classification, etc.)for the source files is insufficient and/or inconsistently maintained (R1) - Source files contain sensitive personal data but misclassified as less sensitive categories, resulting in violations against the "need-to-know" security principle for end users of ChatBot (R2)	- Privacy and security - Misinformation
Chunking and embeddings	- During the chunking process, documents are broken into smaller components. - The embedding process then converts the smaller pieces into high-	After the chunking and embedding processes, raw data is converted into numerical vectors. Any sensitive personal information or business secrets within original files, after the processing, becomes much more difficult to identify, and the original data tagging could be lost or become less relevant. - Lack of proper logical access control to the vector databases (R3) - Vectors may be modified without	- Misinformation - Privacy and Security

	dimensional vectors so an AI system can retrieve information semantically.	proper authorization (i.e., embedding manipulation), leading to output quality issues (R4)	
Indexing and retrieval	- Maintain hybrid search: dense vectors for meaning; sparse indices for rare terms, IDs, and exact phrases. - Apply permission, jurisdiction, and freshness filters before semantic similarity search to narrow the candidate set.	These four steps collectively play a crucial role ensuring the output of the ChatBot meets the expectation of the business user groups (i.e., retrieval quality), and the poor execution would lead to: - Hallucinated output which deviates from the approved artifacts (e.g., policies, procedures, or system manuals, etc.) (R5) - Irrelevant or misleading output leading to poor decision-making of ChatBot users (R6) - Lack of association between system output and validated internal artifacts makes it difficult for end users to verify the validity of the output (R7)	- Misinformation - AI system safety, failures, and limitations
Re-ranking	- Reorder the top-N retrieved candidates with a stronger cross-encoder or reranker trained on in-domain relevance labels to promote truly on-point evidence and demote distractors.		
Context assembly	- Pack evidence into the model's context window with clear structure (titles, sections, snippets) and process near-duplicates. - Favor fewer, higher-quality snippets with explicit citations and effective dates over indiscriminate concatenation.		

Grounded generation	- Use prompts that instruct the model to answer only from provided context, cite or quote sources, and declare uncertainty or scope limits when coverage is insufficient. - Enforce output schemas and post-process for policy adherence including but not limited to safety, security, privacy checks.		
Evaluation and monitoring	Leveraging measurable metrics, proper evaluation ensures the application	Poor evaluation of RAG systems result in user requirements cannot be met, and the risks are captured with the following:	- Misinformation - AI system safety, failures, and limitations
	meets the business requirements in quality, system latency, and development cost, etc.	- Metrics for output quality and system performance are not well understood by the testing team (R8) – Methodology for monitoring & reporting model drift is not properly established (R9)	

3.2. In-depth Assessment and Detailed Planning with 3LOD Model

The development of risk scenarios concluded with 9 risk scenarios associated with the implementation of ChatBot. These risk scenarios are either AI-specific (i.e., R5, R6, R7, R8 and R9) or generic IT/Cyber risks which can be amplified by the AI use case (i.e., R1, R2, R3, and R4).

In order to evaluate the risk ratings for individual scenarios, the project team started by evaluating the application's inherent risk. After reviewing the firm's criteria for mission-critical applications, the consensus was reached that the ChatBot would not be considered as a mission-critical application, and the following factors had been taken into account for the assessment of inherent risk: Recovery Time Objective (RTO), Recovery Point Objective (RPO), data volume, data sensitivity, business processes supported, financial and non-financial impact of system outage and/or malfunction, etc.

Leveraging the AI Risk Tiering model defined by EU AI Act (European Commission, 2025), this internal use case would closely match the profile of "Limited" risk tier. The risk profiling with ChatBot provides Firm X with an excellent opportunity to experiment their 1st AI use case and then get key learnings applied to future use cases while being able to limit the damage in case of control failures.

Considering that the key infrastructure systems are hosted internally and have no public-facing interfaces which can be easily accessed by adversaries, all risk items were subsequently assessed as "medium" level for their residual risk levels, requiring proper mitigating action plans as the firm has no appetite to accept operational risk levels rated above "Low".

3.3. Risk Ownership and Accountability

Having recognized that a formal AI Risk Committee was not in place when the project was being launched, the accountability for identifying and addressing the application's risks (both AI specific and non-AI specific) was assigned to the Head of IT, which is also the project sponsor.

The project, as an experiment for Firm X, aims to provide key learnings for guiding future AI use cases; specifically, it has two main goals to achieve from a risk governance perspective:

1) To identify gaps in existing risk management processes for managing AI risks

2) To propose short-term solutions addressing such gaps as well as longer-term solutions for moving the firm's maturity for AI governance to the next level

Having these goals in mind, the project team proactively engaged the key stakeholders including business user groups and relevant stakeholders to participate into the project from end to end, ensuring concerns and feedback from different teams can be collected and addressed in a timely manner. The risk management activities in this case can be largely grouped into two categories, a) pre-implementation risk assessments, and b) post-implementation assessments & ongoing runtime monitoring.

In the following, we demonstrate how the firm addressed the 9 risks using the existing risk governance processes to the extent possible, and how the gaps were remediated.

R1 – Metadata (e.g., ownership, security classification, etc.) for the source files is insufficient and/or inconsistently maintained

Existing Controls	**Areas of Concern**	**Proposed Action Plans**		
		1st LOD	**2nd LOD**	**3rd LOD**
Data Management Standard, where metadata requirements have been defined, owned by Data Risk Management Team (2nd LOD)	The implementation of Data Management Standard is limited to mission-critical systems/application, and non-critical systems do not have consistent implementation of the Data Management Standard.	The project team will identify in-scope systems supporting the ChatBot (Pre-implementation). Compliance status (for Data Management Standard) will be confirmed by the system/data owners (Pre-implementation).	Based on findings from 1st LOD, update the risk registrar. (Pre - implementation). Identify opportunities for improving the existing standards, if applicable (Pre-implementation). Follow up on action plans when non-compliance issues are identified by 1st LOD (Pre and post-implementation).	Based on findings from 1st and 2nd LODs, update the audit universe as necessary (Post-implementation). Independently verify the effectiveness of action plans by 1st and 2nd LOD. (Post-implementation).
Key learnings from the experiment	Data governance practices, including data quality and lifecycle management, in the age of AI, require an enhanced level of maturity.			

R2 - Source files contain sensitive personal data but misclassified as less-sensitive categories, resulting in violations of "need-to-know" principles for end users of ChatBot.

Existing Controls	**Areas of Concern**	**Proposed Action Plans**		
		1st LOD	**2nd LOD**	**3rd LOD**
Information Classification Standard, owned by Information Security (2nd LOD)	Classification tagging is a responsibility of information owner, and there are no technical tools in place to enforce the implementation.	The project team will identify in-scope systems supporting the ChatBot (Pre-implementation). Compliance status (for Information Classification Standard) will be confirmed by the system/data owners (Pre-implementation).	Same as R1	Same as R1

Key learnings from the experiment	To support the implementation of AI, data discovery tools, which enable efficient searches of personal data and/or business secrets, would become a necessity. The firm's relatively low maturity in data management has significant implications for the selection of RAG architecture design, as the source data containing personal data or business secrets could end up in the vector databases, which are vulnerable to certain types of adversary attacks.

R3 - Lack of proper logical access control to the vector databases.

R4 - Vectors may be modified without proper authorization (i.e., embedding manipulation), leading to output quality issues.

Existing Controls	**Areas of Concern**	**Proposed Action Plans**		
		1st LOD	**2nd LOD**	**3rd LOD**
Logical Access Management Standard is in place, incorporating the well-known security principles (e.g., need-to-know, minimum privilege, etc.). Server hardening baselines, covering commonly used traditional databases (e.g., IBM DB2, etc.) in alignment with CIS Benchmark, have been established within the firm, but vector databases are out of scope. A formal Information Security Risk Assessment (ISRA) is in place, and major IT changes or projects shall go through the ISRA process.	As a net new category of databases for the firm, there are no clear guidance providing instructions on how to classify data contents and/or security hardening requirements. The existing ISRA process, maintained by Information Security, was designed for traditional non-AI applications, and the assessment does not consider the AI-specific risks.	Develop and implement a configuration baseline for vector databases, incorporating best practices from the industry (Pre-implementation). Elaborate the technical, physical and administrative controls for the vector databases (Pre-implementation).	Refresh and execute the Information Security Risk Assessment (ISRA) process to ensure it captures AI-specific risks (Pre-implementation).	Update the audit universe as necessary (Post-implementation). Independently verify the effectiveness of action plans by 1st and 2nd LOD (Post-implementation).
Key learnings from the experiment	Due to the firm's risk-averse nature for data breaches involving personal data, the project team ended up with a very conservative architecture design, which identified and masked (or redacted) sensitive data at the stage of ingestion (shown in Figure 4). This way, the post-embedding information placed in the vector databases contain no personal information or business secrets. Compared with more sophisticated designs (e.g., a role-based access model) which could enable certain end-user groups to access sensitive data when there is a valid scenario (e.g., a select group of end users may be granted access for certain types of business secrets), the final selection of the architecture design was well aligned with the intended business objective as well as the firm's overall maturity level for data management and cybersecurity.			

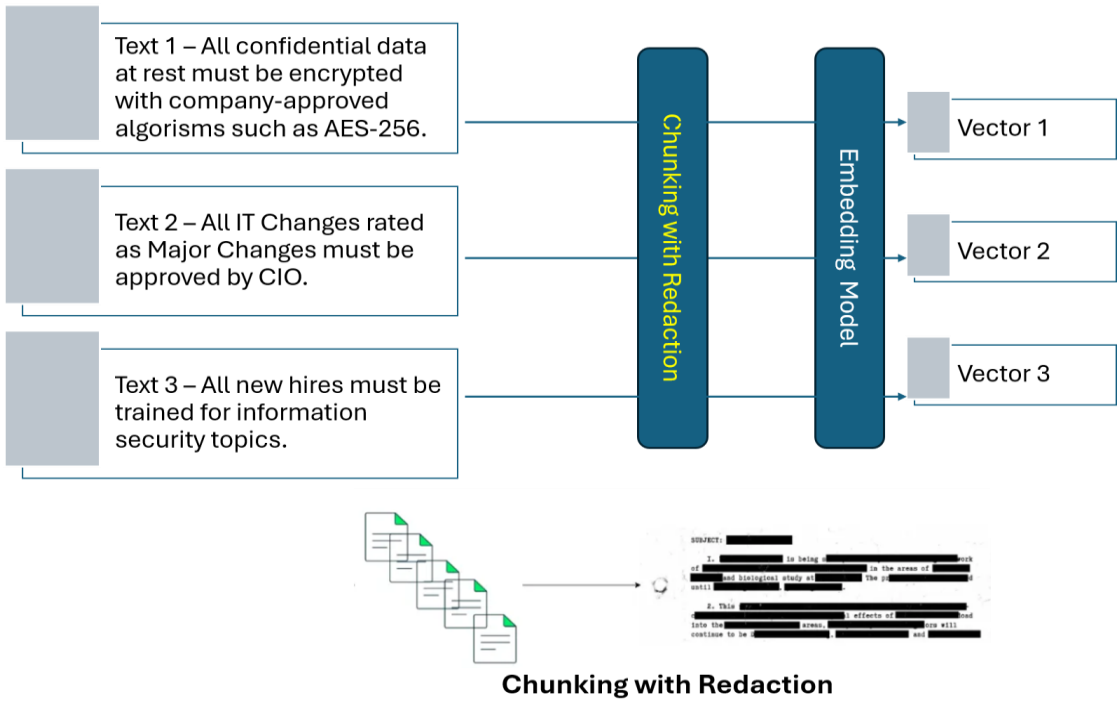


Figure 4. Chunking and Embedding with Sensitive Information Redacted / Masked

While the Chatbot, by design, has the capability of supporting translation, summarization, and document reformatting, the risk analysis elaborated below focuses specifically on the retrieval-based question-answering capability that uses approved internal artifacts as its knowledge sources. Risks associated with standalone translation and summarization capabilities were assessed separately, under the third-party risk management program, and they are outside the scope of this case analysis.

R5 (AI – Hallucination): The system generates claims, interpretations, or citations that are not supported by the approved source artifacts provided in the RAG context.
R6 (AI – Retrieval Relevance): The system retrieves or presents information that is irrelevant, outdated, incomplete, or not applicable to the user's business context, potentially leading to inappropriate decisions.
R7 (AI - Lack of Traceability) - Lack of association between system output and validated internal artifacts makes it difficult for end users to verify the validity of the output
R8 (AI - Insufficient Application Testing) - Metrics for output quality and system performance are not well understood by the testing team
R9 (AI- AI-system Drift Detection & Prevention) – Methodology for monitoring AI-system drift is not properly established.

Existing Controls	**Areas of Concern**	**Proposed Action Plans**		
The firm has its Model Risk Management Standard in place. Model Risk is a 2nd line function, under the Chief Risk Officer. The introduction of new application or major business process is subject to the	The Model Risk Management Standard was primarily designed for traditional statistical, rule-based, or machine-learning models and may not adequately address the end-to-end risks associated with GenAI/RAG systems, including hallucination, retrieval failure,	**1st LOD**	**2nd LOD**	**3rd LOD**
		Pre-implementation: The project team to produce a detailed AI use case inventory to document the AI use case characteristics including but not limited to: version number, target performance metrics, LLM	Pre-implementation: Enhance the Model Risk Review Process to include a risk-based GenAI/RAG assessment covering: intended use and limitations; source-data and retrieval controls; hallucination and relevance risks;	Pre-implementation: Provide independent, non-assurance advisory input on auditability, documentation, evidence-retention, and future audit-scope considerations. Confirm that the proposed governance approach provides sufficient audit trails for decisions,

firm's Change Management process, which requires thorough testing prior to production deployment.	citation traceability, prompt/configuration changes, and vendor-model changes. The existing Change Management Process may not adequately address ongoing AI-system performance monitoring or vendor-initiated changes. GenAI/RAG applications therefore require additional system-level monitoring throughout their lifecycle.	vendor / model number, system recovery targets, application criticality rating, etc. Conduct testing for hallucination rate, retrieval relevance, citation coverage and correctness, response latency, error rates, and user acceptance. Testing should include direct and indirect prompt-injection scenarios. Define minimum acceptance thresholds and documented launch, escalation, and rollback criteria. Post-implementation: Own day-to-day application monitoring, incident management, and remediation. Maintain a dashboard and perform periodic quality reviews using automated monitoring. Escalate when application metrics reach pre-defined thresholds. Key metrics as part of ongoing monitoring: Groundedness / unsupported-claim rate; Hallucination and inaccurate-response rate; Retrieval relevance and failed-retrieval rate; citation coverage, citation correctness, and source freshness; User feedback,	citation traceability; prompt, model, and configuration versioning; testing methodology; acceptance thresholds; residual-risk approval; and ongoing monitoring requirements. Independently challenge the adequacy of the evaluation dataset, testing results, metrics, thresholds, and proposed mitigating controls before production deployment. Post-implementation: Review periodic KRI/KPI reporting and independently challenge deteriorating trends, material incidents, and remediation plans. Assess whether changes require re-testing, formal re-review, risk re-rating, or escalation to the appropriate governance committee. Changes to the LLM vendor/model version, prompt, embedding model, retrieval or reranking logic, vector database, knowledge base, access-control design, or intended use should be assessed against documented materiality criteria. A formal re-review should be required when a proposed or detected change meets the material-	approvals, testing, model/system changes, monitoring, incidents, and remediation. Post-implementation: Update the audit universe and audit plan, as appropriate. Confirm that documentation including model inventory, test outcome, approvals, version records, and monitoring records are retained properly. Independently assess the design and operating effectiveness of 1st and 2nd LOD controls, including AI risk management lifecycle, incident management, change management, escalation protocol, etc.

		complaints, escalation, and task-success rate; latency, error rate, availability, and operational incidents.	change criteria.	
Key learnings from the experiment	RAG-related risks cannot be managed through traditional model validation alone. The firm requires a risk-based AI-system review that covers the end-to-end RAG pipeline, addressing source data, retrieval, generation, citations, configuration changes, and runtime monitoring, etc. For GenAI applications, ongoing monitoring should extend beyond conventional model drift to cover end-to-end AI-system performance, including knowledge-base freshness and coverage, retrieval quality, vendor-model behavior, and changes in user-query patterns. Intentional changes to prompts, configurations, or system components should be governed through change management.			

IV. KEY TAKEAWAYS AND RECOMMENDATIONS

As demonstrated from the above analysis, AI applications such as RAG-based solutions bring their unique risks which cannot be fully addressed by the traditional risk governance model, which relies on risk management processes to cover distinct domains (e.g., data security, privacy, model risk, application operations, etc.). On the other hand, based on the case study, we have demonstrated that AI risk management can be integrated into an organization's existing risk processes with appropriate revision, requiring a coordinated effort supported by the senior management.

When organizations embark on AI adoption journeys, it would be beneficial to start with comprehensive gap analysis on their risk governance model for understanding where the gaps are located. For the unique attributes which are not sufficiently addressed by the current risk processes, organizations should update their processes accordingly before kicking-off the risk review of the AI use case. Although individual organizations could vary greatly in the maturity of risk management and may have very different views on how AI risks should be addressed, the end goals should the same – the AI uses cases must meet the business objectives while the internal policies and external regulations must be followed. To ensure organizations can address AI risks systemically, we recommend that the organizations' existing risk management processes should be reviewed against an industry-accepted AI Risk Taxonomy, such as NIST AI RMF (N. AI, 2024), or MIT AI Risk Taxonomy (Slattery et al., 2026). The integration of AI also means the organizations must re-visit their current Key Risk Indicator (KRI) program and other metrics for application testing and monitoring. With AI technologies move extremely fast, the traditional review intervals for such metric programs may become insufficient, which should be a factor to be considered from a risk monitoring perspective.

Also importantly, organizations should design and maintain their AI risk governance model to make it flexible enough to accommodate the speed of AI innovations. Over time, the robustness and completeness of the risk governance model will evolve as organizations gain more experiences on the journey of AI adoption. Experimenting with non-mission-critical applications, which may be internal-facing and carry low-sensitivity data, would allow the organizations gain valuable experiences while ensuring the operational, legal and compliance risks stay within their risk appetite. The gaps in AI expertise should also be recognized and documented as the human-related limitations need be properly addressed before organizations increase their use cases in complexity or volume.

Last but not the least, the reality is that most organizations cannot maintain all necessary AI expertise in-house, and external third-party services will be inevitable for various tasks and processes. Therefore, it is also important to re-visit the 3rd party risk management practices to ensure the AI risks can be captured properly.

REFERENCES

[1] AI, N., 2024. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST Trustworthy and Responsible AI Gaithersburg, MD, USA.

[2] AI, G., 2024. Guardrails AI | your enterprise AI needs guardrails — guardrailsai.com. URL: https://www.guardrailsai.com/docs/.

[3] BCBS, 2011, Basel III: A global regulatory framework for more resilient banks and banking systems – revised version June 2011, Available at https://www.bis.org/publ/bcbs189.htm

[4] Bilal, A., Ebert, D., Lin, B., 2025. Llms for explainable ai: A comprehensive survey. arXiv preprint arXiv:2504.00125.

[5] Beasley, M., Branson., B., Hancock, B., 2010, How Key Risk Indicators can Sharpen Focus on Emerging Risks, Committee of Sponsoring Organizations of the Treadway Commission.

[6] Vorvoreanu, M., Heger, A., Passi, S., Dhanorkar, S., Kahn, Z., Wang, R., 2023. Responsible AI Maturity Model – Mapping your organization's goals on the path to responsible AI, Microsoft Aether Central UX Research & Education.

[7] BigID, 2025, AI Risk & Readiness in the Enterprise. https://home.bigid.com/download-ai-risk-report. European Union, 2016. General data protection regulation (GDPR) – article 25: Data protection by design and by default. https://gdpr-info.eu/art-25-gdpr/. (Accessed 03 June 2025).

[8] European Commission. (2025, April 9). The AI Continent Action Plan. European Commission. https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan

[9] F5 Cloud Market Trend Report, 2025. RAGs to Riches: Using RAG and AI Data for Enterprise AI. Fang, X., Li, J., Mulchandani, V., Kim, J.-E., 2025. Trustworthy AI on safety, bias, and privacy: A survey. arXiv:2502.10450. URL: https://arxiv.org/abs/2502.10450.

[10] Government of Canada, 2000, Personal Information Protection and Electronic Documents Act, Available at https://laws-lois.justice.gc.ca/eng/acts/p-8.6/.

[11] Moreton, Á., and Jaramillo, A, . 2021. "The problem of complete, irreversible data anonymization." AIhub, July 6. https://aihub.org/2021/07/06/the-problem-of-completeirreversible-data-anonymisation/.

[12] Muria-Tarazon J.C., Oltra-Gutiérrez, J.V., Oltra-Badenes, R., Escobar-Román, S., 2025. "Uncovering Research Trends on Artificial Intelligence Risk Assessment in Businesses: A State-of-the-Art Perspective Using Bibliometric Analysis, Special Issue Emerging Technologies of Human-Computer Interaction, Appl. Sci. 2025, 15(3),1412; https://doi.org/10.3390/app15031412
ISO/IEC, 2021. ISO/IEC TR 24029-1:2021 – Artificial Intelligence (AI) – Assessment of the Robustness of Neural Networks Part 1: Overview. Technical Report, ISO/IEC, Available at https://www.iso.org/standard/77608.html.

[13] International Organization for Standardization, 2023. ISO/IEC 42001:2023 – Artificial Intelligence Management System (AI MS) – Requirements. Technical Report, ISO/IEC, Available at https://www.iso.org/standard/81230.html.

[14] Office of the Superintendent of Financial Institutions, 2026, FIFAI II: AI Risks and Opportunities: Adopting an AGILE Framework in Canadian Financial Services, Available at https://www.osfi-bsif.gc.ca/en/about-osfi/reports-publications/fifai-ii-ai-risks-opportunities-adopting-agile-framework-canadian-financial-services.

[15] Office of the Superintendent of Financial Institutions, 2025, Guideline E-23 for Model Risk Management, Available at https://www.osfi-bsif.gc.ca/.

[16] Office of the Superintendent of Financial Institutions, 2022, Guideline B-13 for Technology and Cyber Risk Management, Available at https://www.osfi-bsif.gc.ca/.

[17] Raza, S., Sapkota, R., Karkee, M., Emmanouilidis, C., 2026, TRiSM for Agentic AI: A review of Trust, Risk, and Security Management in LLM-based Agentic Multi-Agent Systems, AI Open.

[18] Schaar, P., 2010. Privacy by design. Identity Inf. Soc. 3 (2), 267–274.

[19] Slattery, P., Saeri, A., Grundy, E., Graham, J., Noetel, M., Uuk, R., Dao, J., Pour, S., Casper, S., Thompson, N., The AI Risk Repository: A meta-review, database, and taxonomy of risks from artificial intelligence, 2026, Patterns.

[20] The Institute of Internal Auditors, 2020. The IIA's Three Lines Model – An update of the three lines of defense, Available at https://www.theiia.org/globalassets/documents/resources/the-iias-three-lines-model-an-update-of-the-three-lines-of-defense-july-2020/three-lines-model-updated-english.pdf.

[21] Thompson, G., 2023. "Data Anonymization in AI: A Path Towards Ethical Machine Learning." Security & Privacy (blog), November 22. www.privacydynamics.io/post/ data-anonymization-in-ai-a-path-towards-ethical-machine-learning/.

[22] Yin, R.K., 2018, Case Study Research and Applications: Design and Methods, 6th Edition. Sage Publications, Thousand Oaks, CA.

★★★