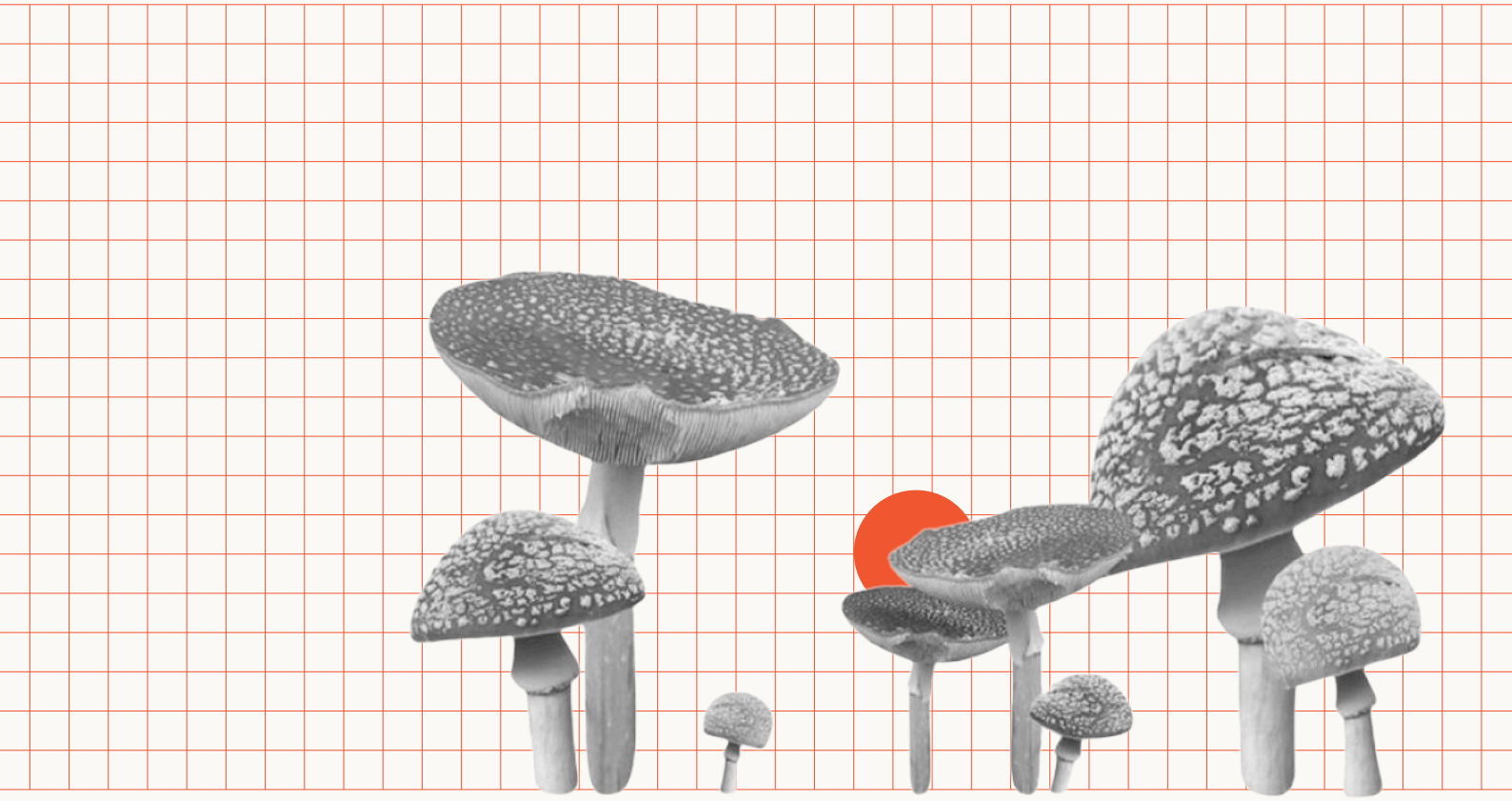
A guide to the emerging AI security risks for Enterprises



CISO Guide

Redactive

Table of Contents

Introducing the Technology Drivers Exposing AI Security Risks

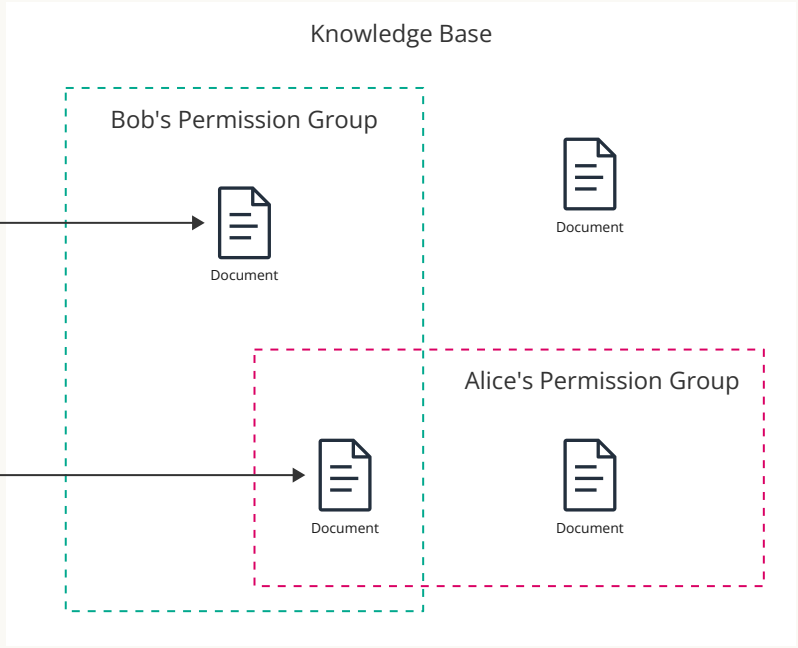
As Artificial Intelligence (AI) evolves, organizations are facing a reshaped threat landscape where traditional security measures fall short. The root of this challenge lies in data sprawl across numerous knowledge bases both within and beyond the company's Virtual Private Cloud (VPC). Managed SaaS tools like Confluence and SharePoint, which emphasize document sharing as a core feature, inadvertently expose organizations to heightened risks. Each document within these platforms requires precise, fine-grained permissions to control who can and cannot access sensitive information.

The task of assigning these permissions often falls to employees who create or inherit the documents, leading to a high potential for human error and unintentional exposure. This model creates a gap that security teams struggle to bridge effectively, especially when managing permissions for thousands of users across sprawling data assets. In this environment, where misconfigurations can expose sensitive data, deploying robust permission management strategies becomes a priority for CISOs to mitigate risks associated with the rapid adoption of AI tools, particularly Generative AI (GenAI), which can process vast amounts of data with unprecedented accuracy.

Understanding How GenAI Interacts with Knowledge Bases

Corporate knowledge bases are often large collections of unstructured data that serve as essential resources for employees. Unstructured data includes documents, images, emails, and videos—types of information that lack rigid categorization, complicating security protocols. Knowledge bases such as Sharepoint and Confluence are the main sources of unstructured data in an organization.

With the adoption of GenAI tools like AmazonQ and Copilot, employees can access and analyze vast amounts of unstructured data more efficiently. These tools leverage semantic search capabilities to deliver accurate, context-rich responses, significantly enhancing productivity and enabling employees to find relevant information quickly and effectively.



In a Knowledge Base users will have access to different documents based on their roles. Sometimes documents overlap in permission groups, this is good and allows for collaboration.

While this capability unlocks significant value for organizations, it also introduces unique security concerns. Unstructured data within knowledge bases operate with document-level permissions, typically set by individual employees. However, these permissions are often optional and lack validation controls for internal content, meaning sensitive information within specific paragraphs or sections may be inadvertently exposed, even if the overall document is restricted.

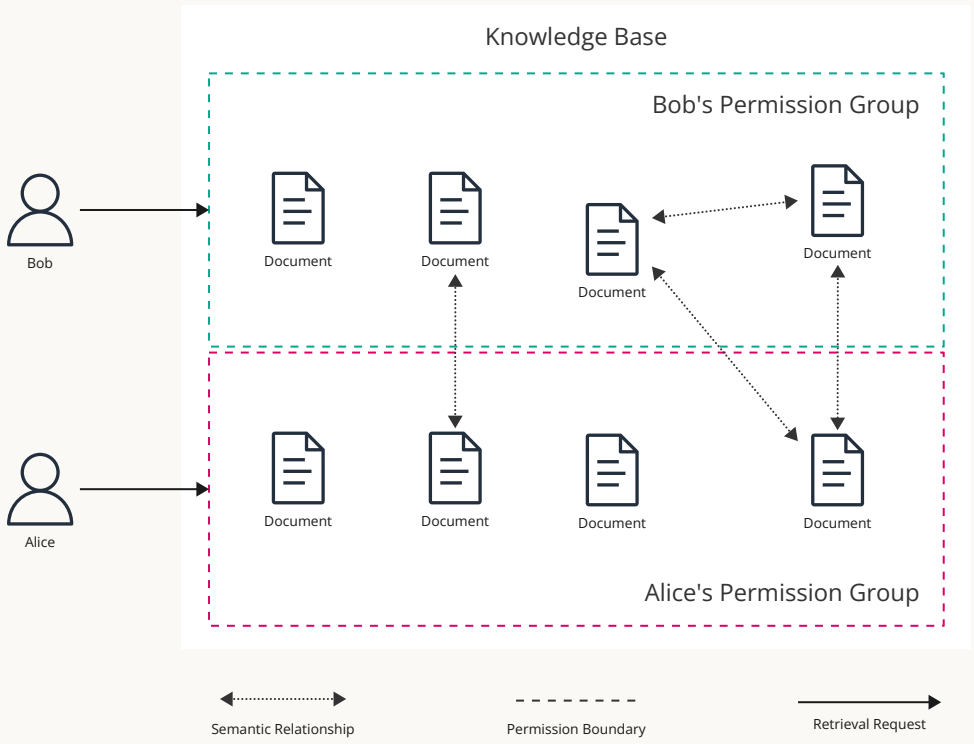
For example, if permissions are set incorrectly, sensitive data could easily be included in AI responses, leading to unauthorized access.

This problem has been experienced at scale by companies such as Salesforce and the Australian Government as they rolled out a Copilot trial across multiple agencies.

> *Their information management in SharePoint is not great which has resulted in end users finding information that they shouldn't have had access to, though this is a governance and data management issue - not a Copilot issue.*
>
> Agency representative in Digital Transformation Agency (DTA) interview for Australian Government trial of Microsoft 365 Copilot [2024].

For IT and Security leaders, this creates a significant challenge: how can they reliably identify sensitive data or detect potential leaks across vast, unstructured data repositories without manually reviewing every document? This limitation underscores the need for advanced tools that can automatically detect and enforce fine-grained security controls within unstructured data.

Given the sheer volume of documents and data sources, coupled with the fact that permissions are often managed at the individual document level by employees, it becomes nearly impossible for CISOs to comprehensively audit and manage permissions at scale. As such, the deployment of GenAI systems demands that organizations implement automated tools to continuously monitor and manage their permission structures.



Permissions and information don't always align. In this example of a badly configured knowledge base some documents have semantic similarities with other documents that different users have access to, meaning that either data is leaking or users don't have the full context required to do their jobs.

GenAI's Impact on the Threat Landscape

As GenAI becomes a standard feature in enterprise systems, its comprehensive understanding of organizational data changes the security landscape in fundamental ways.

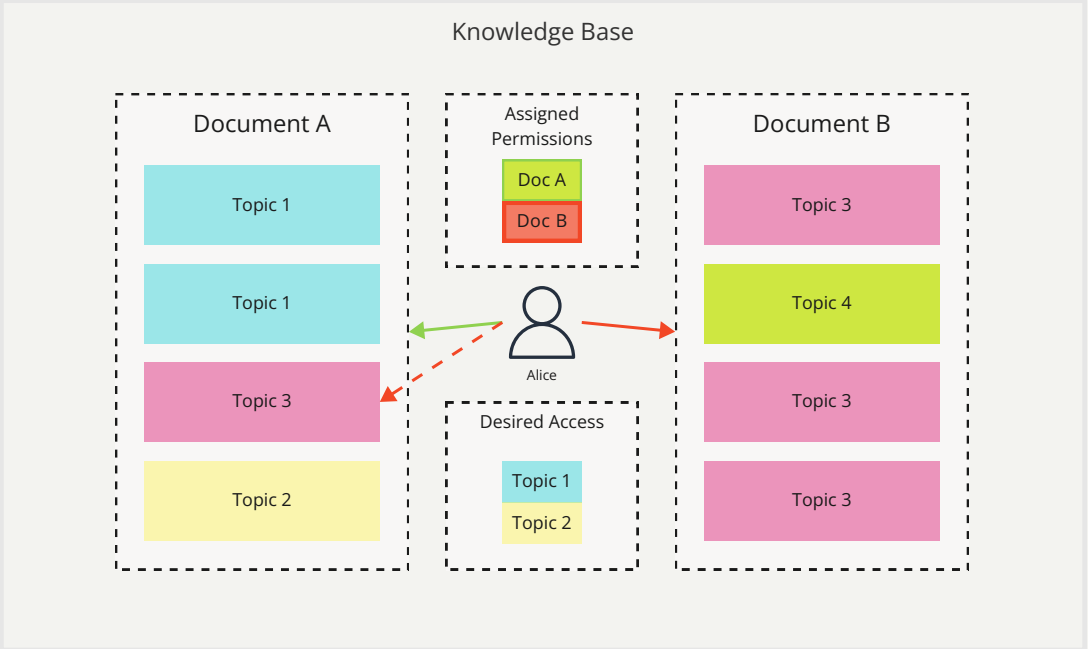
Fine-grained permissions on individual business assets play an essential role in enhancing operational resilience, as they provide precise control over access rights. **Misconfigured permissions, however, remain one of the most prevalent security vulnerabilities within organizations**. Broadly applied permissions can result in employees or unauthorized users accessing confidential data, increasing the likelihood of data breaches.

> *A user (Bob) in accounting can make an innocuous query for "sales reports for last financial year" the response that Bob is expecting would contain extracts from the sales reports over the last year. However with Generative AI and a permission misconfiguration, Bob is also shown another users' (Alice in sales) performance report that shows she is under review due to her poor sales performance in the last financial year. While the root cause of this data leak is a misconfiguration of permissions in the knowledge base, it is Generative AI that causes the leak to be exposed.*

This type of scenario illustrates how even well-intentioned queries can lead to significant security breaches due to GenAI's pervasive access to unstructured data within a knowledge base.

> *... Microsoft (Copilot) lacks the data, metadata, and enterprise security models... That is why Copilot spills corporate data, and forces customers to build their own LLMs. - Marc Benioff*

The introduction of GenAI tools necessitates more precise control over access permissions, as improperly managed data could lead to unauthorized access and widespread data exposure.

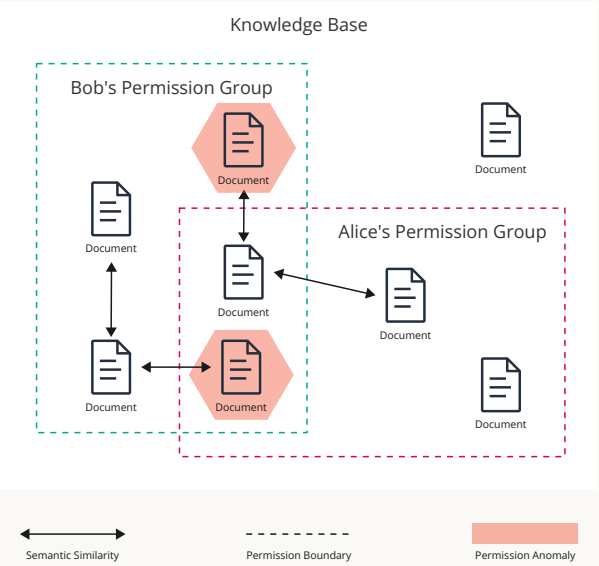


Document permissions often don't align with the desired access controls for a user and can lead to data leaks. For example, Alice should only be able to access documents pertaining to topics 1 and 2 so they have been given read access to document A. Alice is not permitted to see information about topic 3 and has been denied access to documents containing the topic. Due to an oversight document A has sensitive information about topic 3 which Alice can see.

Limitations of Traditional Security Tools in Handling AI and Unstructured Data

Standard Data Security Posture Management (DSPM) tools traditionally relied on non-AI mechanisms, such as keyword detection and predefined rules, to detect unauthorized data access or misuse. These rule-based systems are often configured to identify patterns associated with sensitive information like personal identifiers, financial data, or confidential terms. Current DSPM frameworks in the market do not address the risk that GenAI with semantic retrieval presents, specifically it doesn't resolve the core misconfiguration of data assets risks that exist.

AI Data Security tools provide an advanced solution by understanding the document's meaning, allowing for a nuanced approach to data security. AI Data Security tools can analyze the semantic structure of documents, capturing the intent behind information rather than just identifying specific words. AI Data Security can identify sensitive information within paragraphs or contextually link related content across different documents, offering a comprehensive view of potential security risks above what DLP can identify alone. This capability enables AI Data Security tools to catch subtle or emerging threats, allowing for more proactive security.



AI Data Security Tools can identify when documents are misconfigured based on the information found within. In this image 2 documents identified in red have been detected as having anomalous permissions based on how other documents are related semantically in the knowledge base.

Incomplete Permissions Table

	Document A	Document B	Document C	Document D
Bob	✓	✓	✗	?
Alice	✗	✓	?	✓
Susan	✗	✓	✗	?
Hasan	✓	?	?	✓
Elijah	✗	✓	?	✓

AI Data Security tools can detect anomalies in document permissions by using techniques such as collaborative filtering to understand the permission boundaries inside of a Knowledge Base and recommend appropriate corrections to secure and uplift the contents.

While traditional governance platforms like Microsoft Purview classify and monitor data based on keywords and metadata, these systems are limited by their reliance on static rules. Redactive's AI Data Security Agent, on the other hand, fills critical gaps in data security by leveraging AI to analyze context and intent. **For CISOs, AI Data Security tools are becoming essential in building resilience against GenAI-induced vulnerabilities, as they provide the contextual awareness necessary to address complex data breaches and insider threats.** Redactive is the flagship AI Data Security tool that identifies and resolves the risks found in your data when scaling GenAI tools across your business.

For CISOs preparing to implement GenAI tools, prioritizing the identification and remediation of permission misconfigurations is essential. With Redactive's advanced capabilities organizations can detect, analyze, and address permission vulnerabilities across data sources, reducing risks and enhancing data governance. Redactive scales to handle the large size of enterprise knowledge bases where data is spread across different sources-of-truth like emails, Sharepoint, Confluence, and Microsoft Teams. Redactive helps secure sensitive data in real-time by integrating permission-aware indexes, which minimizes data exposure risks, supports compliance, and streamlines data management.

Recommendations for CISOs: Strengthening Data Security for AI-Integrated Systems

Conduct a Knowledge Base Audit

To build a secure environment for AI-driven data systems, CISOs should begin with a comprehensive audit of existing permissions. This audit should identify any misconfigured or overly broad permissions that could potentially expose sensitive information. Securing and uplifting knowledge bases with an accurate permission model is essential, as it limits unauthorized access by ensuring users have only the access necessary for their roles and tasks. Redactive specialised in AI Data Security with advanced auditing capabilities for knowledge bases such as Sharepoint and Confluence.

Enable AI Data Security Agents to proactively manage risks

Given the unique challenges introduced by GenAI, deploying tools that provide context-aware permissions is essential. For instance, AI Data Security solutions can detect and monitor unauthorized access based on the context of user queries and data content. Additionally, establishing continuous monitoring protocols is crucial, particularly as organizations scale AI deployments. CISOs should consider automated tools to monitor and flag permission anomalies in real time, especially in environments where human error may lead to inadvertent permission changes.

Educate and Prepare Employees

Training employees on best practices for permission hygiene is also vital. Many misconfigurations arise from simple user errors, which can be mitigated through regular training on secure data handling. Given the rapid pace of AI adoption, security teams need to establish and enforce data governance policies that address the unique risks AI presents.

Create a GenAI Risk Management Framework

As AI tools continue to reshape enterprise operations, CISOs should ensure that security is integral to the AI deployment roadmap. This includes establishing a GenAI risk management framework, which outlines governance policies and compliance standards to be maintained as AI use grows. By integrating security protocols into AI rollout plans, organizations can address potential vulnerabilities proactively.

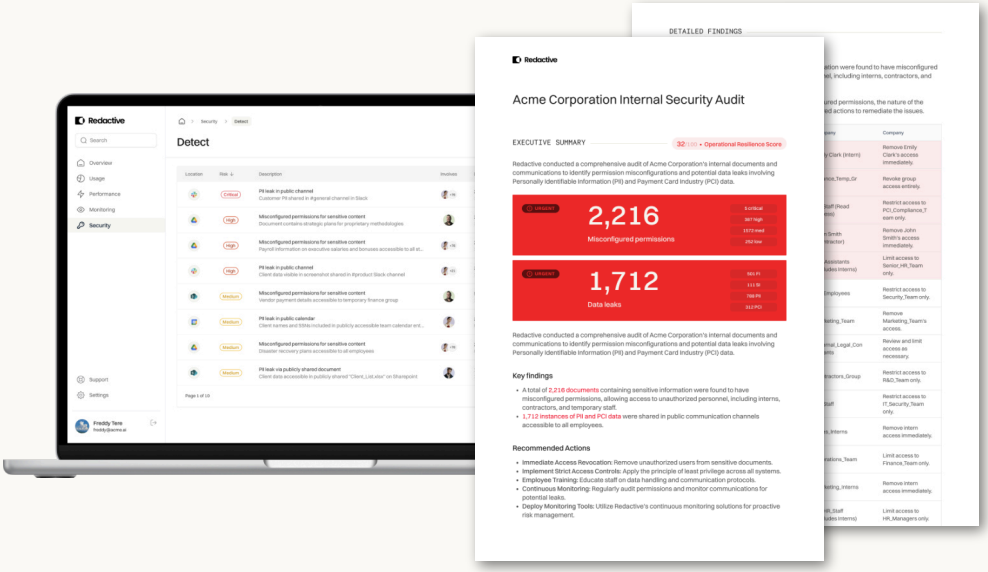
As more products enable Generative AI, maintaining security through fine-grained permissions will be crucial. Security leaders must rethink traditional data governance strategies to ensure these technologies enhance, rather than compromise, operational resilience.

By focusing on robust permission management, continuous auditing, and adopting semantic-aware tools, organizations can strengthen their defenses against emerging AI threats. The proactive security measures CISOs implement today will be critical for safeguarding enterprise data in an evolving digital landscape.

CISOs must prioritize foundational security measures, including fine-grained permissions, dynamic access controls, and GenAI-specific governance frameworks, to mitigate the risks associated with AI-driven insights.

Book a free consultation.

Find out how to effectively use AI in your organisation.



Redactive

Redactive is the trusted AI Data Security platform for enterprises. By understanding the semantic content (meaning) of your operational data, we can reduce your risk posture and prepare banks, super funds, and large corporations to confidently adopt AI tools and navigate data stewardship regulations such as GDPR, ISO27001, and CPS230 at scale.

Redactive.ai's Permission Assurance accelerates regulated organisations to secure & remediate misconfigured overshared permissions across their knowledge bases, preparing it to be used by AI applications such as AmazonQ, Copilot and Glean or your custom AI tools built specifically for your workflows.

Find out how to tackle data leaks ahead of your AI tools exposing them

Book your consultation here

Andrew Pankevicius, Co-Founder
apankevicius@redactive.ai

Alex Valente, Co-Founder
avalente@redactive.ai