CYERA

WHITE PAPER

Securing Generative AI Applications Built on Amazon Bedrock with Cyera

How Cyera's Data-First Approach Delivers Visibility, Governance, and Protection for Generative AI and Agents

WRITTEN BY

Yuri Duchovny, Director of Solution Architecture, Cyera

Executive Summary

From startups to global enterprises, companies everywhere are finding new ways to bring AI into their work, whether it's driven by executive-driven strategic initiatives or inspired by employees seeking productivity gains. For AI applications to reach their full potential, they must access and utilize the right data, much of which is sensitive or proprietary. However, security teams frequently lack visibility into where AI interacts with that data, and how to enforce consistent controls across multiple types of AI: such as widely used public AI services, copilots and embedded AI capabilities within SaaS applications, or custom-built AI systems. The rise of agentic AI, autonomous, non-human agents capable of initiating actions, connecting to systems, and manipulating data without direct oversight, further compounds this challenge. These agents can traverse environments, invoke APIs, and act on sensitive information at machine speed, expanding the attack surface and making data governance, visibility, and protection more complex than ever. As a result, many AI initiatives stall before achieving meaningful scale or business value.

This paper explores how Cyera addresses these challenges by delivering 360° visibility into the AI applications used across an organization and the data they access, together with robust security controls to govern and protect those applications. Cyera AI Guardian builds on the data intelligence powered by Cyera Data Security Posture Management (DSPM) and Cyera Omni Data Loss Prevention, extending their capabilities into AI Security Posture Management (AI-SPM) and runtime guardrails with AI Protect, ensuring data stays secure wherever it interacts with AI.

The OWASP (Open Web Application Security Project) [1] Top 10 for LLMs framework is quickly becoming a go-to resource for organizations looking to understand and mitigate the most common security risks in AI systems. As AI adoption accelerates, this framework provides a much-needed structure for addressing threats such as data leakage, prompt injection, model poisoning, etc... helping teams build safer, more resilient AI applications.

The Amazon Bedrock platform has made it easier than ever to build and deploy powerful LLM-powered applications. With just a few clicks, teams can integrate generative AI into their workflows and products. However, as accessibility increases, so does the risk of unintentionally exposing sensitive data or introducing vulnerabilities.

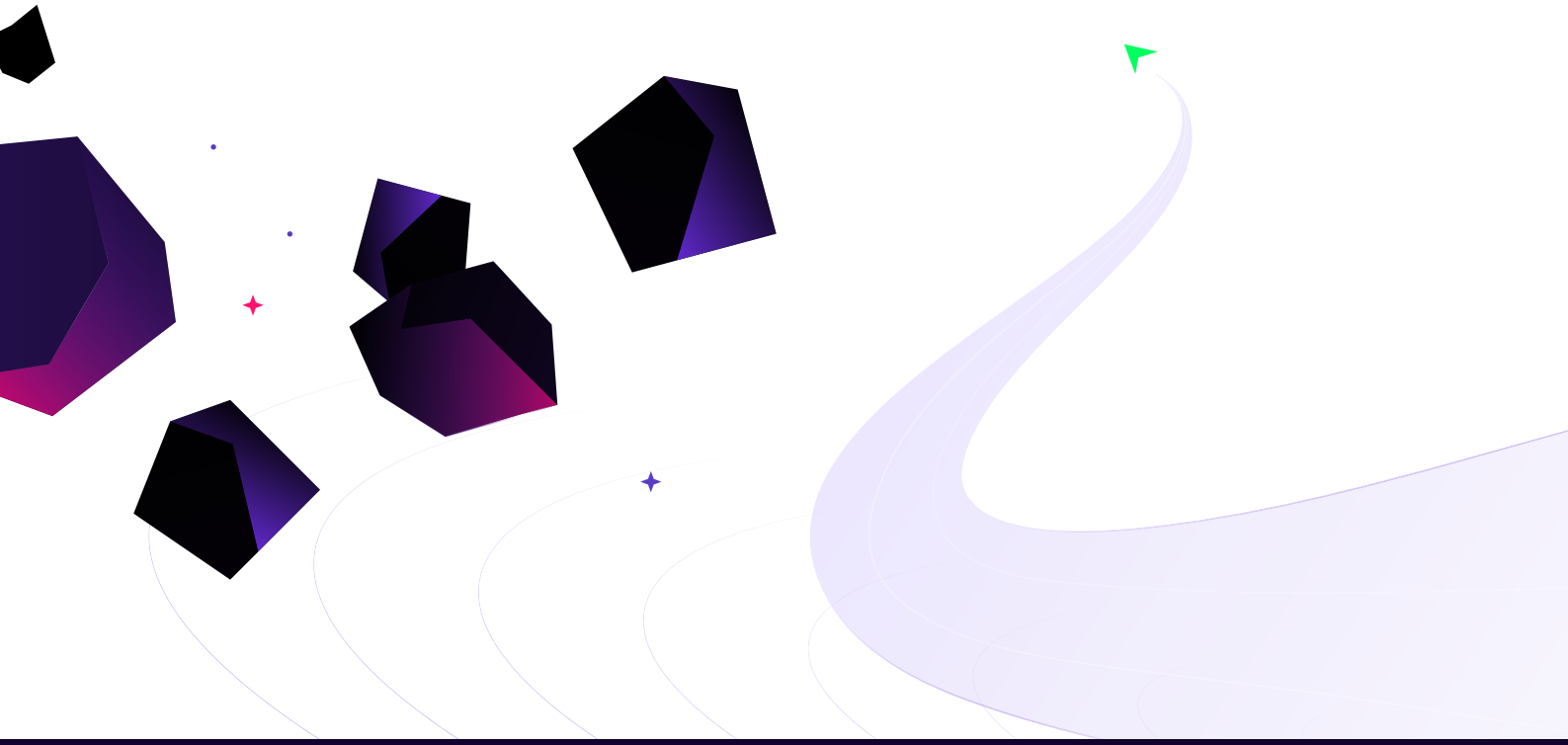
This paper explores how Cyera AI Guardian helps organizations stay ahead of these risks by providing AI-native, data-centric visibility, governance, and protection, including AI Security Posture Management and runtime controls. These capabilities align with the OWASP Top 10 for LLMs framework and help mitigate vulnerabilities such as prompt injection, training data poisoning, and sensitive information disclosure, ensuring that AI innovation remains both fast and secure.

What is Amazon Bedrock

Amazon Bedrock is a fully managed service that provides access to a selection of high-performing foundation models (FMs) from leading AI innovators, including AI21 Labs, Anthropic, Cohere, Meta, Mistral AI, Stability AI, and Amazon itself, all accessible through a single API. It offers the essential tools and capabilities needed to build generative AI applications with a focus on security, privacy, and responsible AI.

What is Cyera?

Cyera is a unified, AI-native data security platform that discovers, classifies, governs, and protects sensitive data across cloud, SaaS and on-premises environments. For AI use cases, Cyera extends these strengths with data-centric AI Security Posture Management (AI-SPM) and inline runtime controls (AI Protect) to keep sensitive data out of prompts, completions, knowledge bases, training sets, and agent tool calls, and to protect agents against prompt injection and related runtime abuses.



Cyera DSPM, Cyera Omni DLP and Cyera AI Guardian Capabilities

Cyera DSPM is the foundational component of Cyera's Data Security Platform. It provides a continuous inventory of data stores across cloud and SaaS environments (e.g., object storage, databases, collaboration platforms) and delivers accurate, AI-native, context-aware classification of sensitive data, reducing classification false positives and enabling focus on real risk.

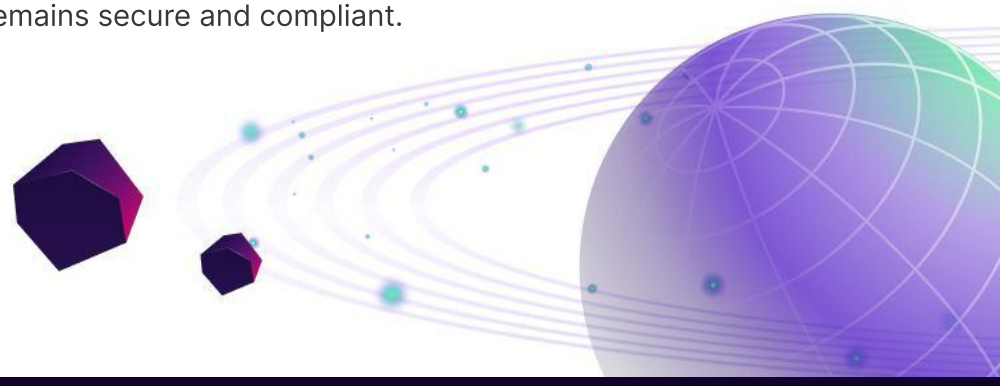
Cyera Omni DLP, built on Cyera DSPM, combines deep, context-aware classification with real-time analysis to protect data in motion. It consolidates alerts from your existing DLP tools and uses AI to prioritize what matters most, explaining why actions were flagged and surfacing related activity. Omni DLP delivers insights to reduce false positives, recommends policy improvements, and focuses attention on high-risk users, data, channels, and third parties, enabling precise, policy-driven protection and continuous optimization.

Cyera AI Guardian unites AI-SPM, which provides a 360° view of all AI assets across the organization (including those built on Amazon Bedrock), with AI Protect, which secures data as it flows into and out of LLMs, agents, and knowledge bases. AI-SPM also surfaces **Shadow AI** across clouds and SaaS, flagging unregistered Bedrock agents, models, and knowledge bases, and mapping them to the sensitive data and identities they touch. In addition, AI-SPM detects Bedrock agents **without configured guardrails** (e.g., content filters, policy checks, safety limits) and raises posture alerts to close those gaps. Together, AI-SPM and AI Protect provide comprehensive visibility and protection across the AI lifecycle.

Built on Cyera's AI-powered DSPM intelligence, AI Guardian establishes the foundation for AI safety controls and governance workflows to remediate potential exposure, over-provisioning, and shadow data growth. It addresses the key patterns in how organizations use Amazon Bedrock to power and scale their AI applications, identifying where enterprise data carries the highest risk and delivering the visibility and control needed to maintain security, compliance, and data integrity. These patterns include:

- **Knowledge Bases for Retrieval-Augmented Generation (RAG)**
- **Model Customization** (retraining or fine-tuning)
- **Agentic Actions (Tools)** connecting to downstream systems

These use cases represent the primary data interaction points and correspond to the **OWASP Top 10 for LLM risks** described above. Cyera AI Guardian is purpose-built to address these challenges. In the following sections, we'll explore how Cyera provides visibility, governance, and protection across each of these patterns to ensure that Bedrock-based AI innovation remains secure and compliant.



Strengthening Amazon Bedrock Security Posture with Cyera AI Guardian

Cyera AI-SPM delivers the first critical value by providing a visibility and governance plane over Bedrock usage across your AWS environment. AI-SPM begins by **inventorying Bedrock assets** (models, knowledge bases, agents) and presents a comprehensive view of what developers are building and using, **streamlining approval workflows** (clear owners, purpose, guardrail status, and policy posture for sign-off) and **surfacing Shadow AI** by flagging unapproved assets so they can be brought under governance.

Within this view, Cyera stitches the ecosystem together by showing which agent is using which model, which draws from a knowledge base, which in turn is backed by a specific data store. It also surfaces rich metadata such as ownership, environment, region, and guardrail status, so approvers see not just what exists, but how it behaves and whether it meets standards. Importantly, each asset carries an AI-powered Business Purpose description, giving admins clear context on how it's being used and, therefore, how it should be secured and governed.

Below is a screenshot from the platform showing this comprehensive view for all AI assets used across the organization, including those built on Amazon Bedrock.

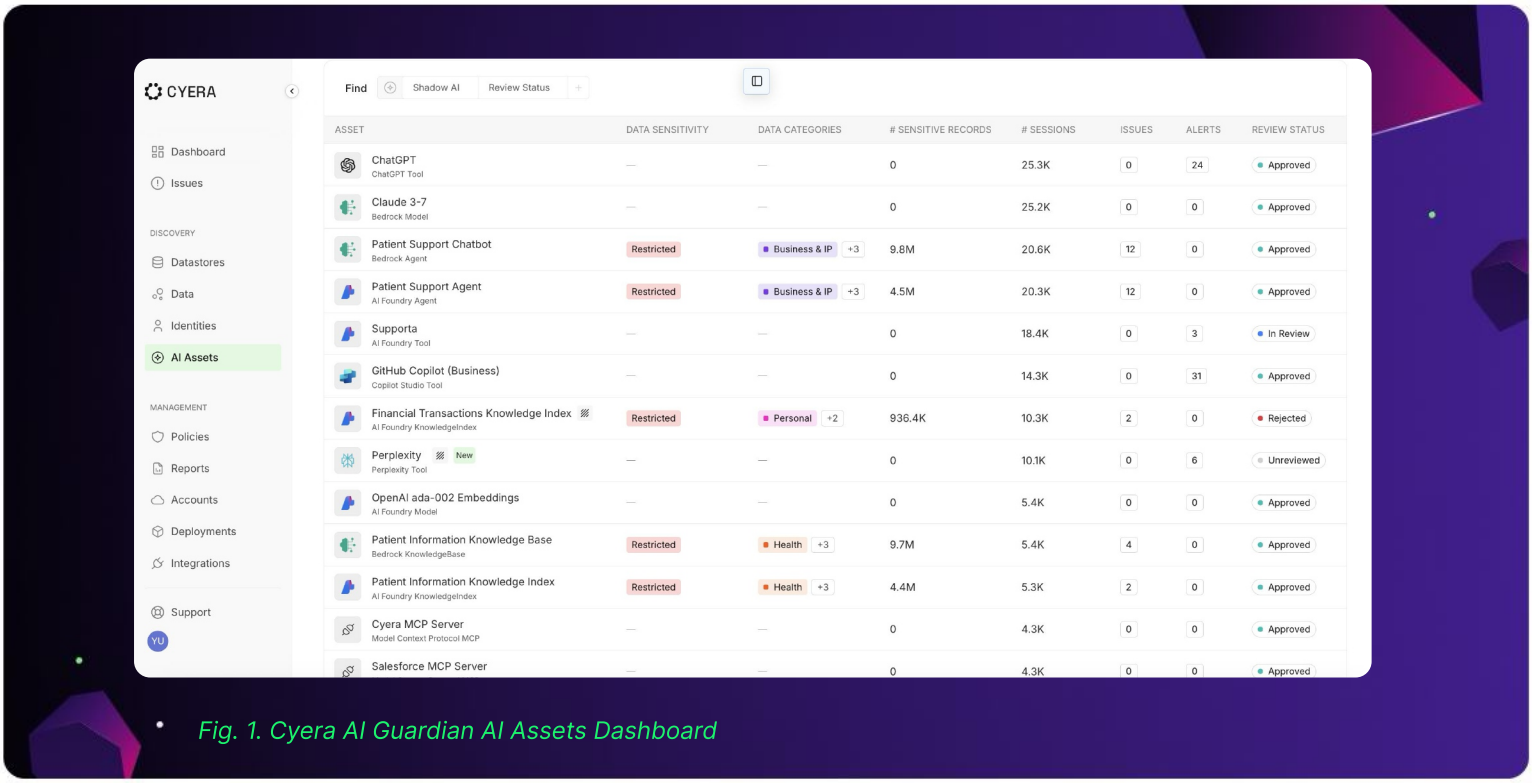


Fig. 1. Cyera AI Guardian AI Assets Dashboard

With this context-rich view in place, we can turn to the specific governance challenges teams face on Amazon Bedrock and how to address them in practice. Organizations are primarily concerned with several business and technical scenarios related to Amazon Bedrock. In the following section, we'll take a closer look at each scenario and demonstrate how Cyera AI Guardian mitigates the associated risks through its unified visibility, governance, and protection capabilities.

Scenario 1: Establishing Company Knowledge Base for Retrieval-Augmented Generation (RAG)

On Amazon Bedrock, retrieval-augmented generation (RAG) grounds model responses in your organization's latest, approved content by fetching relevant passages from a curated corpus (such as S3-backed knowledge bases) and injecting that context into the prompt at runtime. The purpose is to deliver accurate, current, and explainable answers without the cost, latency, and operational overhead of fine-tuning, reducing hallucinations while keeping data in its system of record.

Bedrock Knowledge Bases provide a fully managed RAG workflow, including ingestion, retrieval, prompt augmentation, session context, and source attribution/citations. This functionality brings some potential risks of inadvertently exposing company data. The following issues require mitigation:

Sensitive Information Disclosure:

- Semantic Knowledge Base ingests sensitive data that the end user is not authorized to access, and the agent subsequently retrieves that data.
- A developer mistakenly connects a dev/test agent's Knowledge Base to a production data source containing sensitive information.
- A threat actor adds a plausible but malicious doc to the KB source, steering the agent's behavior and influencing responses.
- Unregistered or test knowledge bases created outside approved processes and connect to production data (shadow AI).

Vector and Embedding Weaknesses:

- A threat actor poisons knowledge base data to manipulate retrievals and influence downstream model behavior.

With Cyera AI-SPM, organizations can mitigate these risks by clearly identifying existing AI assets, such as knowledge bases, LLM models, enhancing this inventory with detailed data classifications within those knowledge bases, and determining which identities have access to the associated data stores.

Because Cyera AI-SPM provides context into what data your AI is accessing and interacting with, it enables teams to prioritize what matters most by surfacing the most sensitive or overexposed data first, and to take actionable steps to mitigate risk by enforcing least privilege, remediating exposure, and monitoring continuously to maintain security and compliance over time.

In addition, Cyera AI Protect enables the implementation of guardrails that filter both queries and retrieved matches from knowledge bases, preventing sensitive data from being returned to the model and exposed to end users.

The high-level architecture diagram for this scenario is presented below:

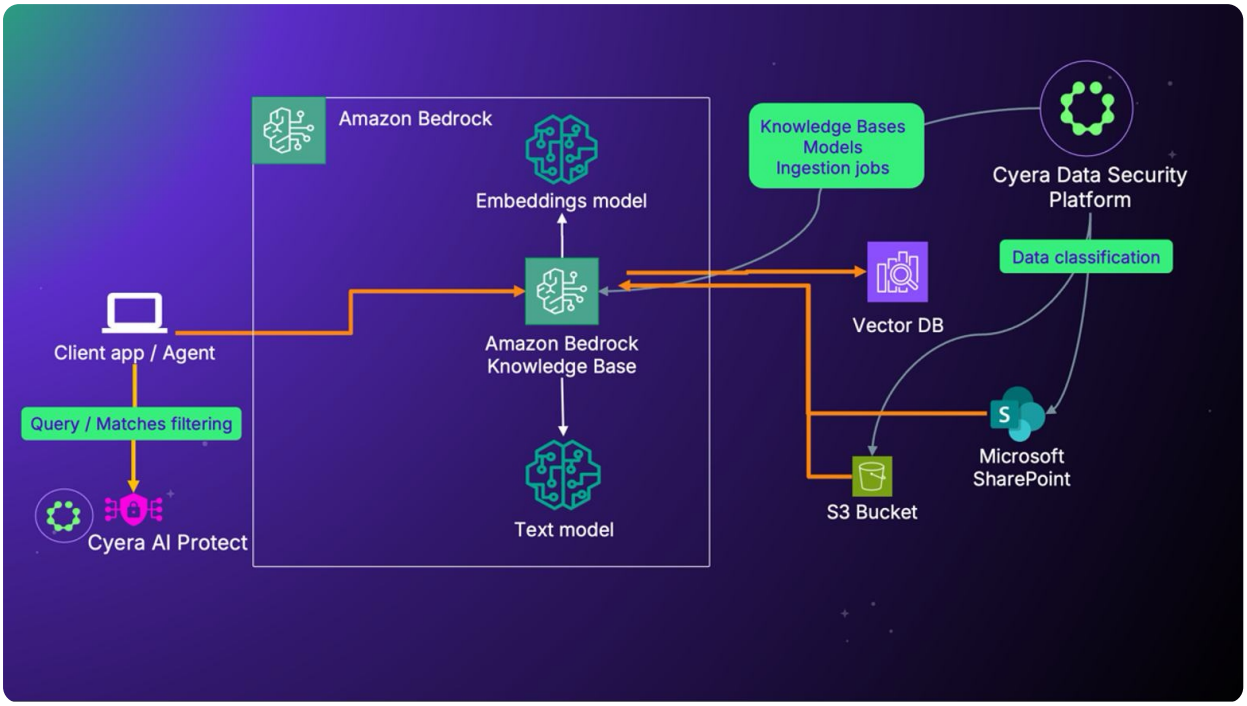


Fig. 2. Cyera AI Guardian for Amazon Bedrock Knowledge Bases

The product images below illustrate how easily administrators can gain a unified view of all AI-related resources, and most importantly, the data contained within those knowledge bases. Cyera provides immediate visibility into security and compliance violations, enabling teams to quickly investigate and remediate issues to maintain a secure and compliant AI environment.

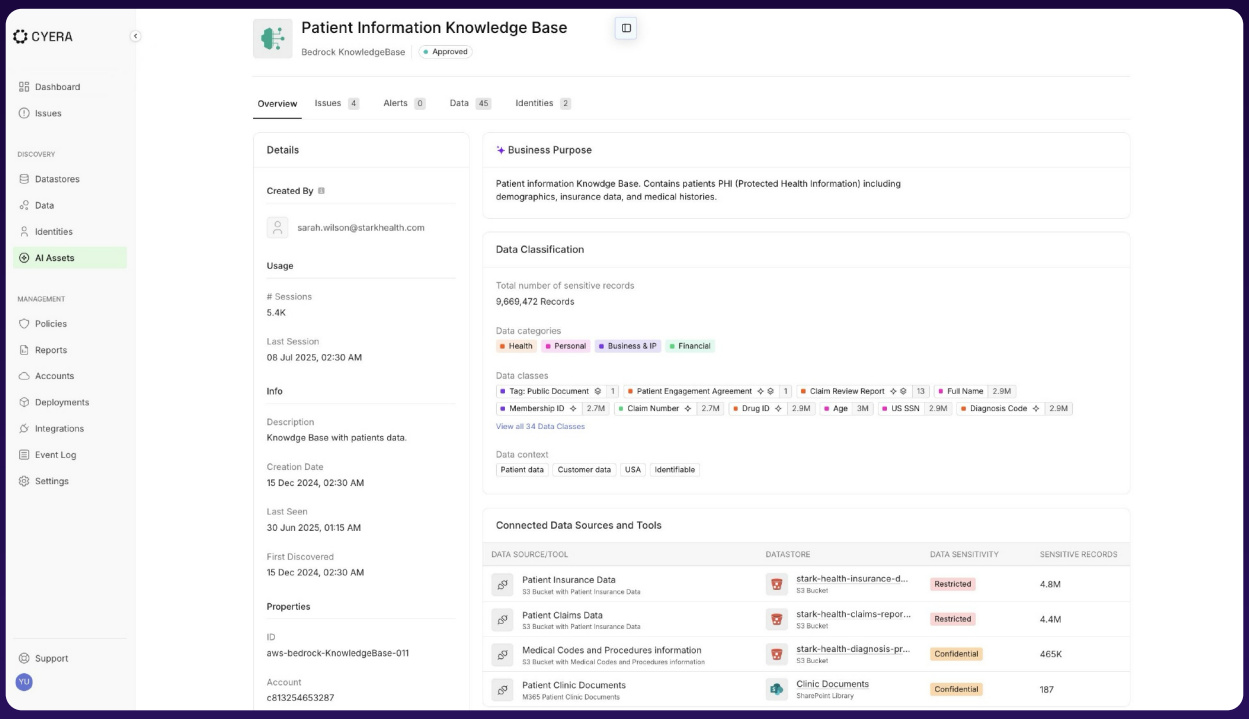


Fig. 3. Cyera AI-SPM - Knowledge Base Details

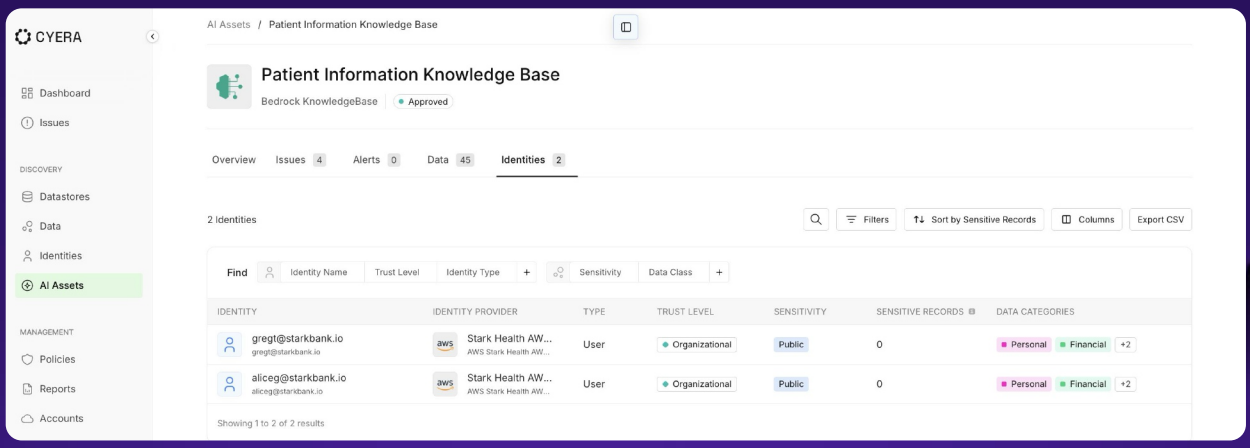


Fig. 4. Cyera AI-SPM - Knowledge Base Identities

Cyera also determines security and compliance issues related to the data in the knowledge base and indicates which records trigger the issues:

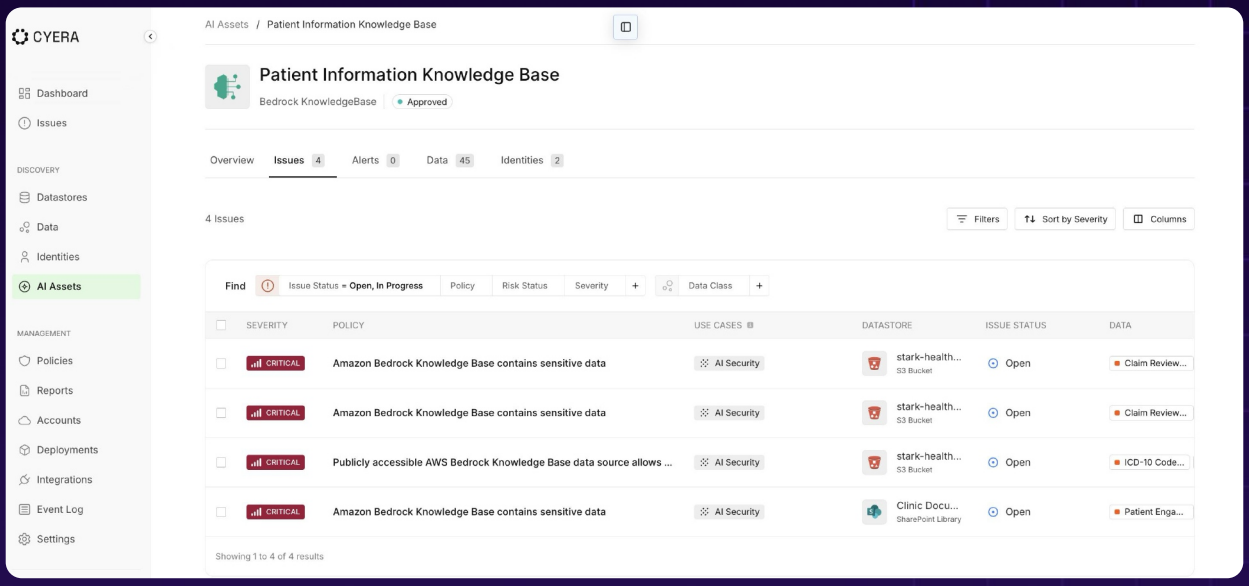


Fig. 5. Cyera AI-SPM - Knowledge Base Issues

Scenario 2: Model Customization

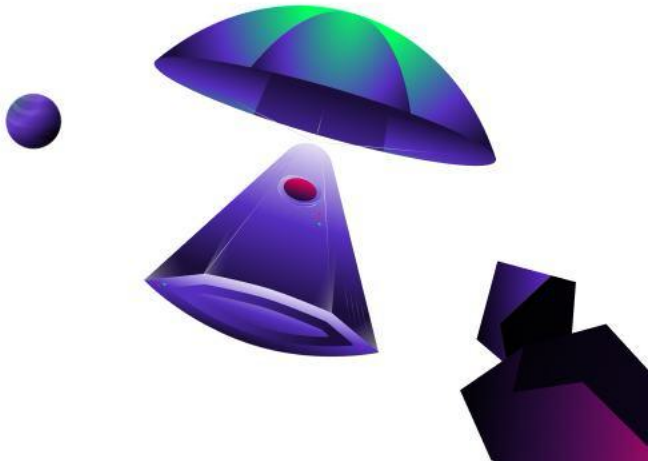
On Amazon Bedrock, model customization and fine-tuning can teach a foundation model your organization's vocabulary, style, and task patterns so that it performs specific jobs with higher accuracy and less prompt engineering. By training on approved examples such as tickets, chats, procedures, you encode stable know-how (formatting, tone, decision rules) directly into the model, which can shorten prompts, reduce tokens, and lower latency and cost at inference. In practice, customization complements RAG: fine-tuning bakes in how to answer, while RAG supplies the latest facts, together delivering responses that are both on-brand and up to date.

In this scenario the main security concerns can be identified as following:

Data and Model Poisoning:

- A model is fine-tuned on sensitive data; due to memorization/overfitting, it may expose that data in its outputs.
- A threat actor injects a crafted dataset into the training/fine-tuning data source to poison the model and influence its behavior.
- Unapproved fine-tuning jobs or custom models running outside governance, trained on sensitive or tainted data.

Cyera AI-SPM automatically discovers custom LLMs deployed on Amazon Bedrock, maps each model to its training and fine-tuning data stores, and flags those that contain sensitive data as potential sources of leakage. It also identifies over-permissive access to training repositories and the identities or IAM roles with unnecessary rights, enabling least-privilege remediation and continuous monitoring.



In addition, Cyera AI Protect enforces guardrails that inspect and filter both prompts and completions at inference time, preventing sensitive data from being returned by the custom model and exposed to end users.

The high-level architecture diagram for this scenario is presented below:

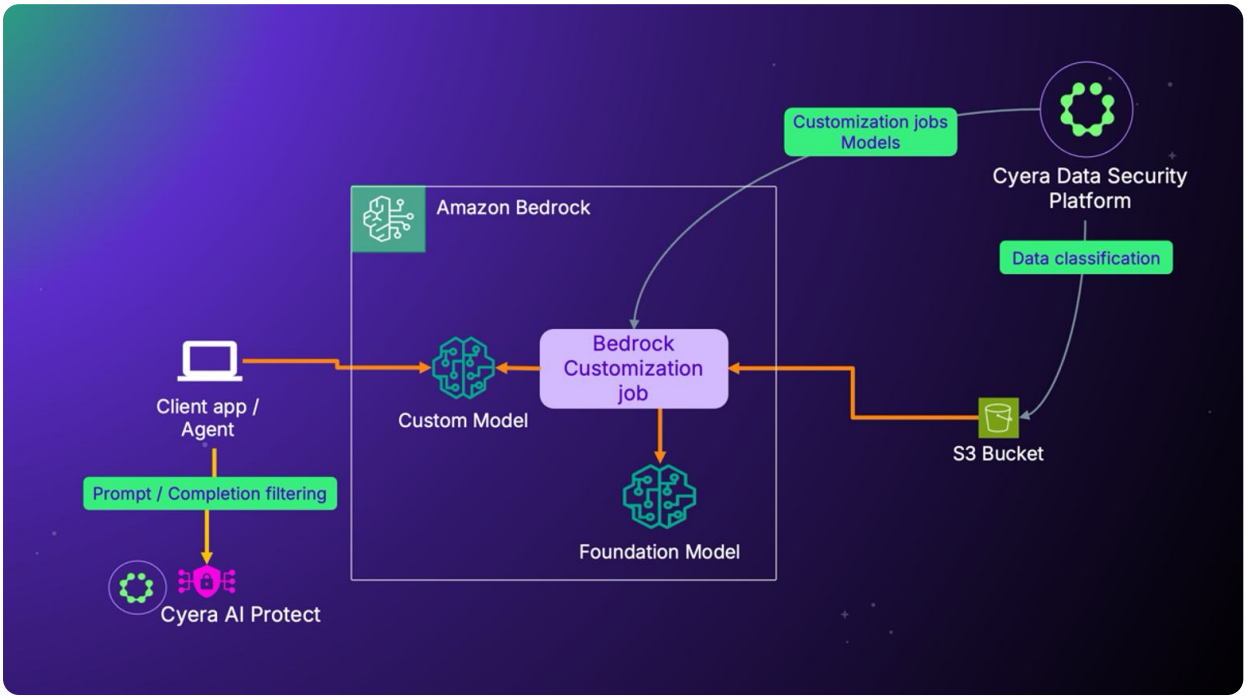


Fig. 6. Cyera AI Guardian for Model Customization

The image below shows how easily administrators can gain a unified view of all AI models, including custom variants, along with their training datasets and the rich data intelligence about the information used to train the model, powered by Cyera.

Cyera provides immediate visibility into security and violations (e.g., sensitive training data, over-permissive access to the training data or the model itself, unapproved models), enabling teams to quickly investigate and remediate issues to maintain a secure and compliant AI environment.

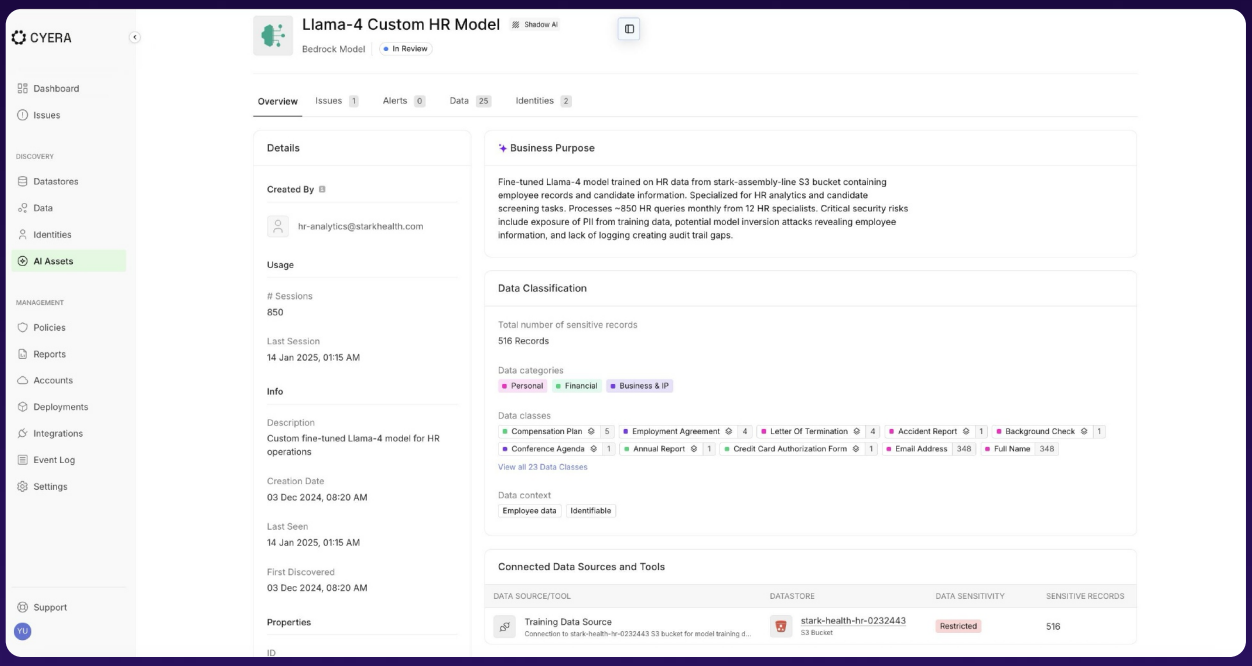


Fig. 7. Cyera AI SPM - Custom Model Details

Scenario 3: Custom Amazon Bedrock AI Agents

Custom Amazon Bedrock Agents transform natural-language intent into auditable actions by orchestrating models, tools (Action Groups/Lambda), and knowledge bases in a single managed runtime. Their purpose is to handle multi-step workflows, gathering context, calling APIs or databases, enforcing schemas, and returning grounded results, without teams having to manually code complex prompt chains or highly proprietary backends.

In addition to the vulnerabilities described above for Knowledge Bases and Custom Models (since agents may use both), additional potential security issues in this scenario include:

Sensitive Information Disclosure:

- An agent in a development or test environment connects to production data sources and retrieves sensitive data in real time.
- Agents connect to unapproved MCP (Model Context Protocol) servers, enabling inappropriate data access beyond sanctioned controls.
- Agent tools retrieve sensitive data without proper authorization on behalf of the end user (missing user context, scopes, or approvals).
- Developers create agents outside approved governance and controls, bypassing review, logging, and security baselines.
- Content and filters are not configured or enforced by the agent.
- Developers create unregistered Bedrock agents or attach new action groups / Knowledge Bases outside approved governance, bypassing review, logging, and baseline controls (shadow AI).
- Developers mistakenly connect irrelevant data to the agent, mismatching its business purpose with its access, violating the least privilege principle and risking that data get exposed to the end user

System Prompt Leakage:

- API secrets or security guardrail logic embedded in the system prompt can be revealed to a threat actor via prompt leakage or debugging outputs.

Prompt Injection:

- Malicious or crafted inputs (including retrieved content) override system instructions or jailbreak the agent.

As we can see, two distinct patterns are associated with this scenario.

First, the goal is to discover and analyze all existing Amazon Bedrock agents, identifying which models, knowledge bases, tools (including MCP servers), and data stores they can access. Cyera AI-SPM performs a detailed inventory of Bedrock agents and correlates them with classified data sources, tools, and identities in use, building on Cyera DSPM data intelligence.

Second, implement real-time protection so that when an agent retrieves data on a user's behalf, it returns only what policy permits. Cyera AI Protect API can be embedded at multiple integration points within agentic applications. On every agent invocation, it can classify and enforce policies on user prompts and agent responses. Beyond that, it evaluates knowledge-base matches, tool calls, retrievals, and completions against organizational policies and data sensitivity, then allows, redacts, masks, or blocks as needed in real time. All decisions are audited, reinforcing least-privilege and continuous compliance.

Below is the architecture diagram that describes this scenario:

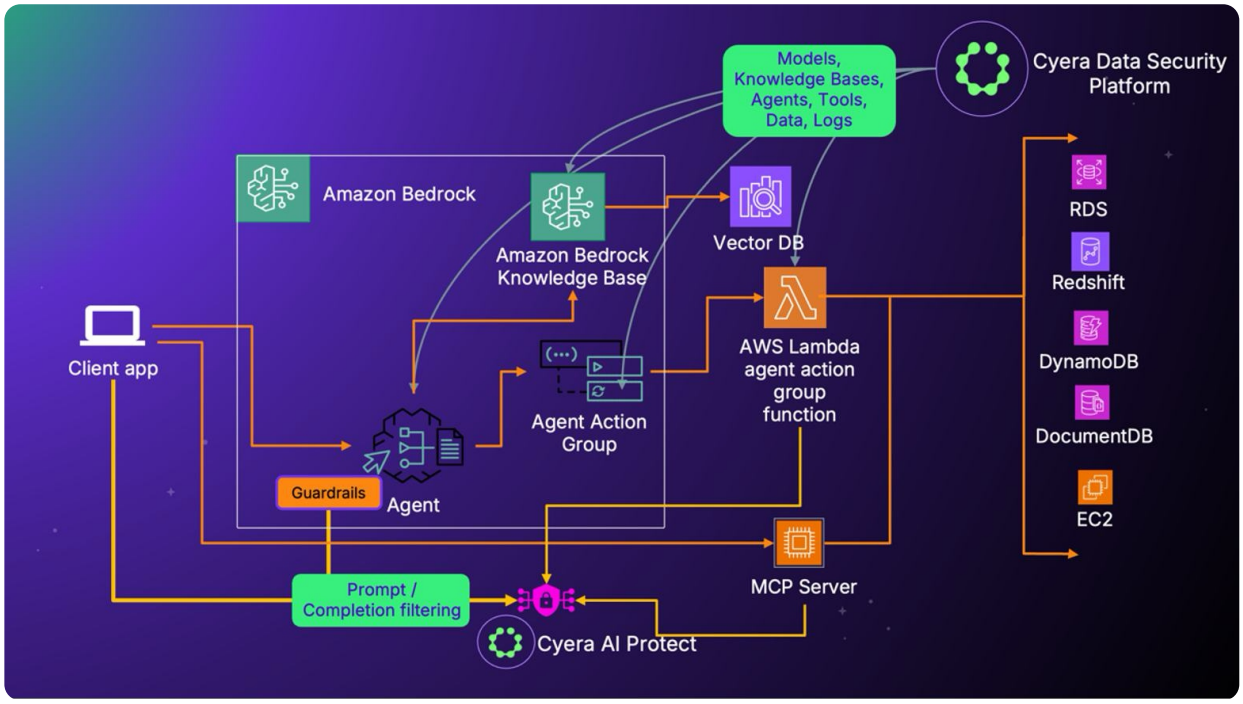


Fig. 8. Cyera AI Guardian for Bedrock Agents

The image below shows how comprehensive Cyera AI-SPM's view of Amazon Bedrock agents is, combining all the moving parts (models, knowledge bases, tools, data stores, identities) with Cyera DSPM data intelligence, policies, and robust remediation capabilities.

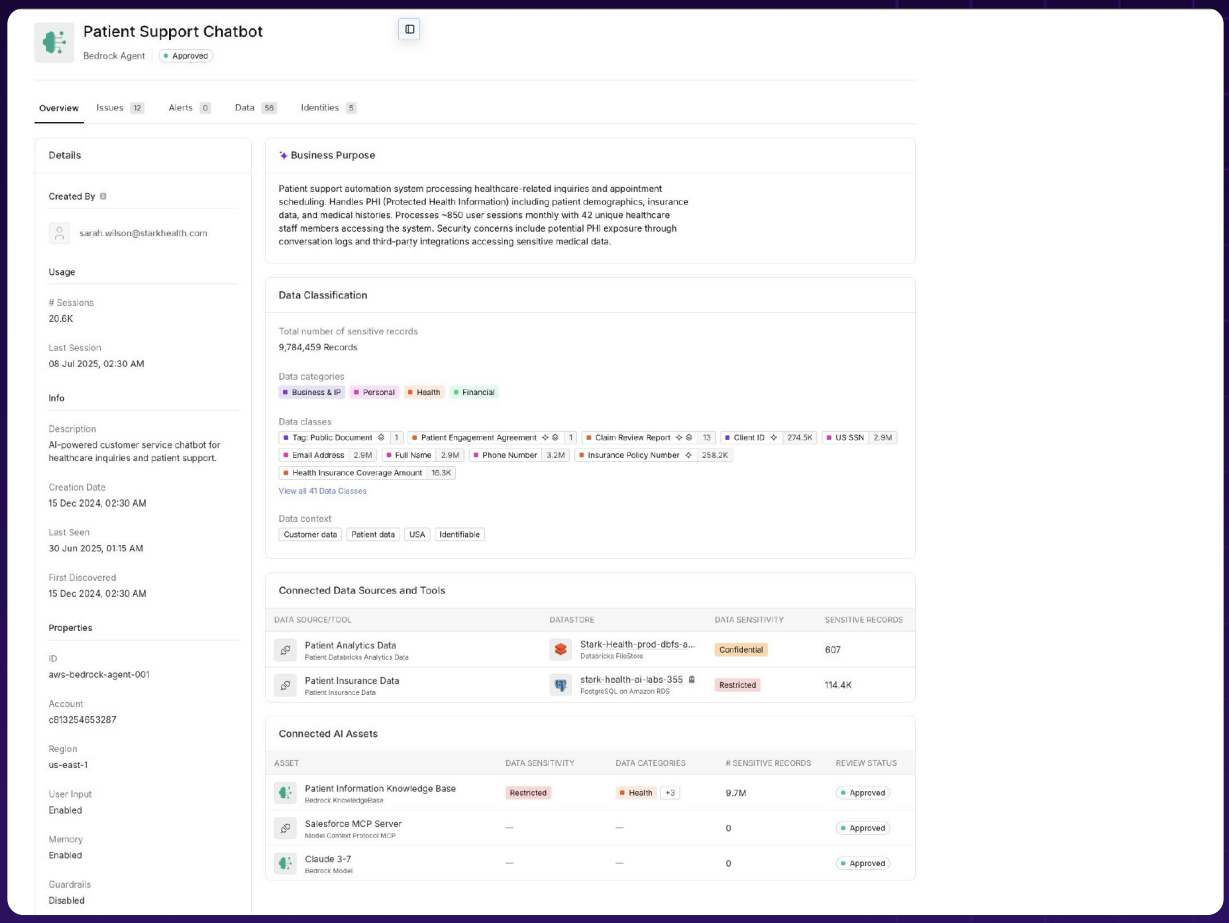
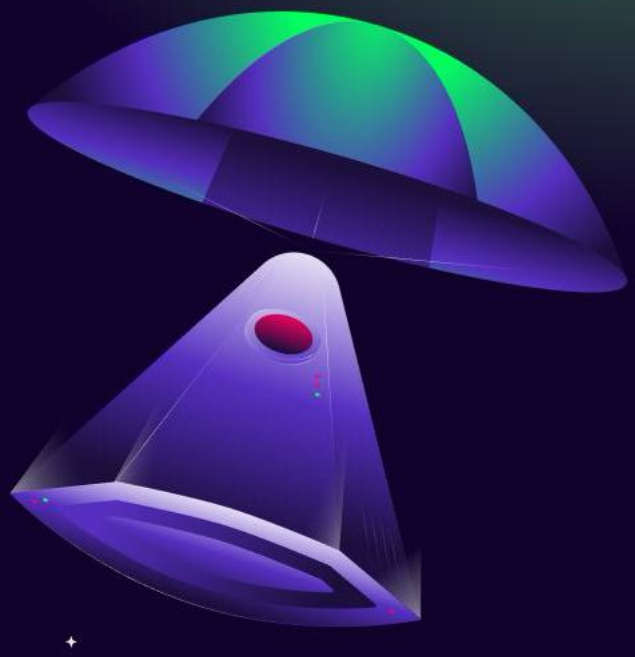


Fig. 9. Cyera AI SPM - Bedrock Agent Details

Conclusion



Generative AI succeeds when it can safely work with the data that makes your business unique. Throughout this paper, we showed that the highest-value Bedrock patterns curating knowledge bases for RAG, customizing models to encode institutional know-how, and empowering agents to act, are also the points where risk concentrates. Cyera's data-first approach resolves that tension. By pairing DSPM's precise understanding of where sensitive information lives with AI-SPM's end-to-end inventory and lineage of AI assets, and AI Protect's real-time guardrails on prompts, retrievals, tool calls, and completions, organizations gain the visibility, governance, and enforcement needed to prioritize what matters, enforce least privilege, and monitor continuously, so Bedrock-based innovation remains secure, compliant, and fast.

A critical part of this is eliminating Shadow AI. AI-SPM's live inventory exposes unmanaged Bedrock agents, custom models, and knowledge bases, while AI Protect contains their blast radius with runtime guardrails.

The path forward is straightforward: connect Cyera to establish a trusted data inventory, onboard Bedrock applications to illuminate access and behavior, and enable runtime protection to keep prompts, retrievals, and tool outputs within policy. Done this way, AI innovation accelerates, compliance posture strengthens, and every Bedrock workload becomes both explainable and auditable by design.

Learn more about Cyera and AWS:
https://www.cyera.io/partnership/aws

CYERA.IO