CYERA

WHITE PAPER

Securing AI Starts with Data Security

Building a Trusted Foundation with Cyera AI Guardian

AI is fueling the next technological era, disrupting almost every industry and shaping the future plans and investments of every organization. Leaders are expected to use AI to foster innovation, create new opportunities, and capture additional revenue streams. Like previous breakthroughs – web, mobile, and cloud computing – many companies will falter in their AI adoption and ultimately disappear, while those that can scale AI effectively and securely will see exponential returns. While AI accelerates innovation, it also brings new security challenges, making safe and responsible adoption critical.

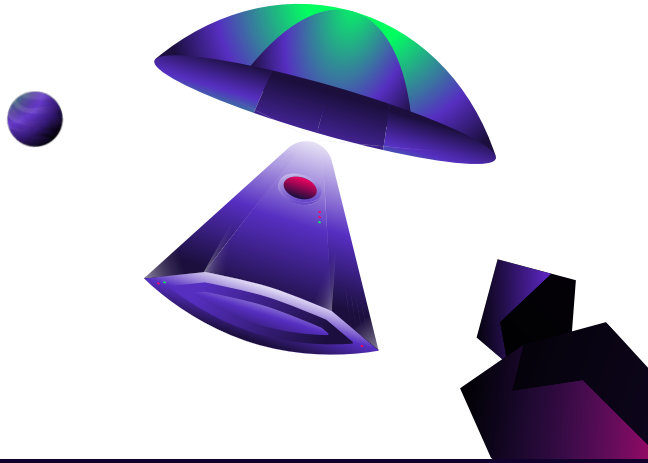
The Balancing Act

Security and innovation are often perceived as a tradeoff versus something that can occur in unison. It's a constant struggle in enterprises, especially with new technologies like AI, where new infrastructure, workflows, and applications are being tested and adopted at unprecedented speed. As a result, many employees bypass security teams as they chase productivity hacks, revenue potential, and anything that will improve time to market. This is especially prevalent in public AI tools like ChatGPT and Perplexity, where sensitive data may be shared and leaked.

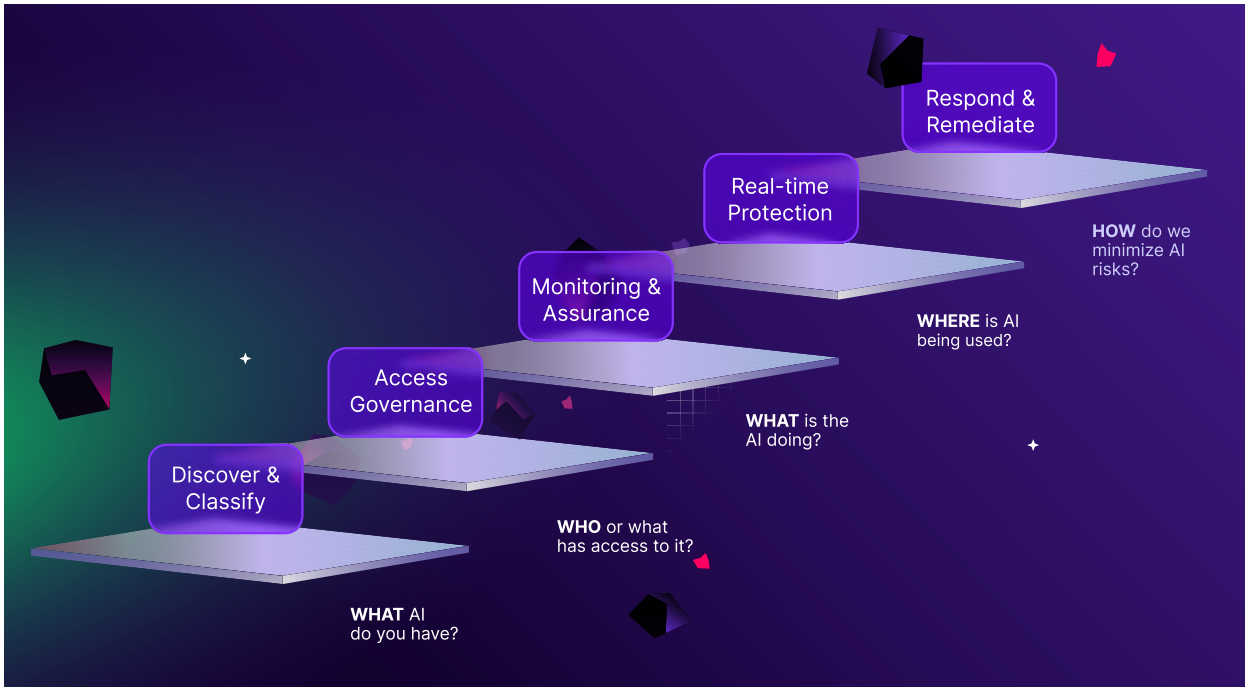
The immense potential of AI has overshadowed conversations around cyber risk

Business leaders must have an AI strategy to stay competitive and security leaders need to empower the business to adopt AI at scale. Organizations that ignore AI adoption risk losing employees, customers, and relevance in the market. But they can't take a laissez-faire or rigid approach as it's the early innings of AI. Organizations need the flexibility to experiment with different flavors of AI and determine which suits their workflows the best.

That's where security can be a strategic driver behind the business initiatives of AI. With the right AI security in place, organizations gain visibility into AI usage without indiscriminate blocking. This empowers teams to not only experiment with and learn AI, but also ensure they have the right monitoring capabilities in place and adjust the guardrails if needed.



The AI Security Journey



We've developed an AI security journey to help guide organizations through the process for adopting AI securely and at scale.

01 **Discovery and Classification:** This initial step involves understanding and categorizing all AI assets within the organization.

02 **Access Governance:** It is crucial to establish robust access controls, ensuring that only authorized personnel and systems interact with AI responsibly.

03 **Monitoring and Assurance:** With proper visibility, organizations can then track AI behavior to ensure compliance and optimal performance.

04 **Real-time Protection:** This involves implementing immediate safeguards to control how and where AI is being utilized.

05 **Respond and Remediate:** Ultimately, teams must be prepared to address and mitigate risks promptly as new threats or exposures emerge.

Together, these steps provide a structured approach to securing AI across its lifecycle.

AI Guardian: Data Security Built for the AI Era

The promise of AI is built upon its unprecedented scale and speed, as well as the emergence of new workflows grounded in new standards, such as MCP and RAG, and semi-autonomous and autonomous agents. Cybersecurity tools need to keep pace with these new advancements in AI. Cyera AI Guardian provides purpose-built security and protection tooling for the AI era.

Building on its foundations in data security, governance, and risk management, Cyera is uniquely positioned to help organizations enter the AI era with confidence — with a unified platform to safeguard how data, identities, and models are used.

AI Guardian combines AI-SPM (AI Security Posture Management) and AI Runtime Protection to deliver visibility into AI assets and proactive defense against emerging threats. Because it's natively integrated with Cyera's DSPM and Omni DLP, it brings the same qualities customers rely on today—rapid discovery of new assets and risks, the ability to operate across the largest enterprise environments without friction, and accurate detection that cuts through noise to highlight the issues that truly matter.

AI Guardian helps organizations answer fundamental questions like:

What types of AI do I have in my environment?

What datastores is my AI accessing?

What identities – human, machine, and agentic – are accessing the AI?

Beyond visibility, AI Guardian helps detect and respond to AI-specific threats. In addition, it helps prevent attacks like prompt injection and data misuse in real-time with Cyera's endpoint API for homegrown AI applications. Finally, it provides the traceability and mapping needed to scale AI safely and respond to audits and other regulatory requirements.

Visibility, Insights, and Risk Assessment with AI-SPM

Visibility is the first step towards a more strengthened (AI) security posture. That's where AI-SPM shines, providing both **breadth and depth** into AI deployments and shadow AI usage. The core capabilities of AI-SPM are **inventory, business insights, and security status.**

Inventory

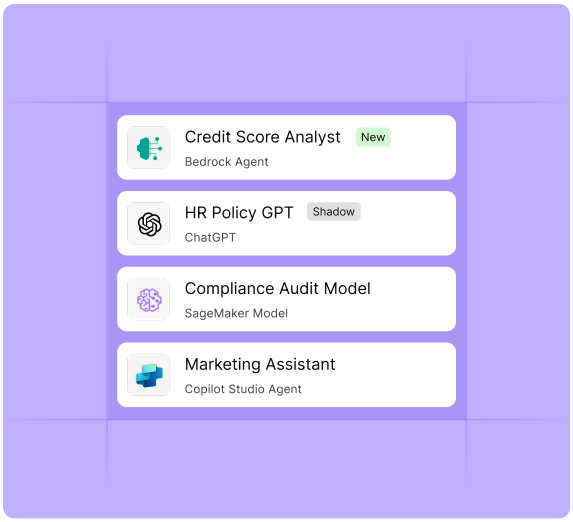
Establishing a complete inventory of all AI systems in use is a foundation of AI-SPM. You can't secure what you don't know exists. AI introduces blind spots where sensitive data, privileged identities, and critical workflows can be exposed or tampered with. Cyera AI-SPM discovers different types of AI:

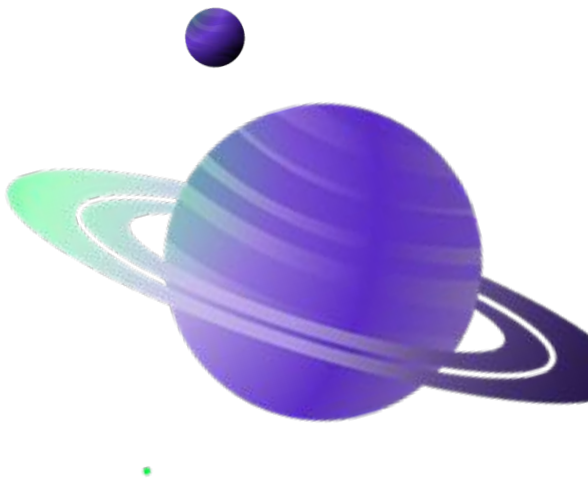
Public AI: Popular tools like ChatGPT, Claude, Gemini and Perplexity. Public AI tools tend to be the main area where shadow AI usage and sensitive data leaks occur. It's also where access controls are the weakest and most difficult to apply.

Embedded AI: Licensed tools like Microsoft 365 Copilot, Gemini Workspace, and Salesforce Agentforce may be "approved", but they still require security controls to enforce the correct handling or interactions with sensitive data.

Homegrown AI: In-house, custom-built AI applications on services like Amazon Bedrock, Azure AI Foundry, and Google Vertex. Ensuring AI tools built in-house are acting as intended is key.

Cyera AI-SPM also flags what AI in use is shadow AI or unapproved, while showing what the most popular AI tools are by user sessions. This not only provides valuable security insights, but also a comprehensive view of organizations' AI deployments.





Business Insights

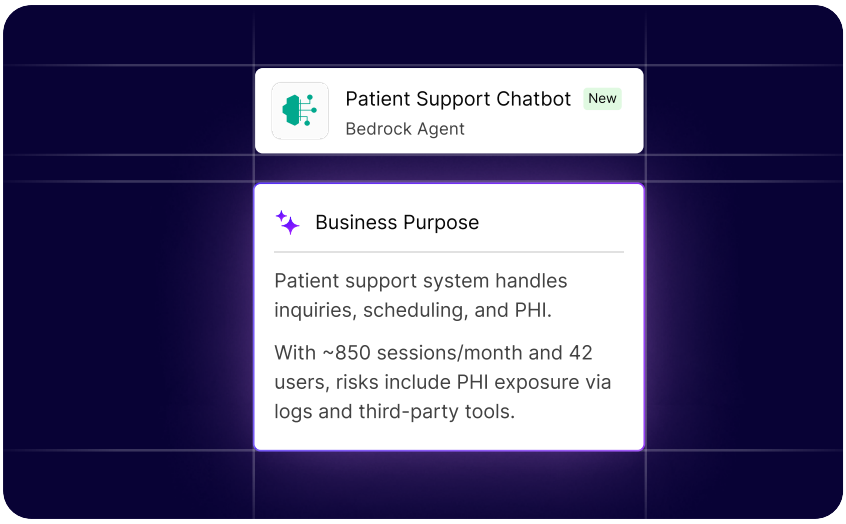
While inventory provides the breadth into an AI security posture, it's the business insights and data mapping that make Cyera stand apart from other AI-SPM vendors. Cyera provides granular details around the type of AI and how it's being used in an organization. Here are the different categories of business insights covered by AI-SPM:

Business Purpose: An AI-generated summary of the business purpose of any AI asset using different contextual clues. For example, a summary for GitHub Copilot could be, "Used by 4000 users for real-time code completion, generation, and refactoring support. Can access sensitive IP and PII data."

Data Sensitivity & Categories: Understand at a high level how sensitive the data is that AI is interacting with, and what categories. Data category examples include: Business & IP, Personal, Health, and Financial.

Review Status: It's likely that IT and security receive requests to approve new AI tools every week, and AI-SPM can help by giving a look into what's approved, in review, rejected, or, in the case of shadow AI usage, unreviewed.

Business insights empower security teams with a wider context to understand what to focus on, providing them visibility into what tools are seeing the most usage and their importance to the business. This sort of context is critical to ensuring that security is working with the business versus 'blocking' it.



OWASP Top 10 for LLMs and GenAI | Recommend Controls

01 **Prompt Injection |**
Prompt filtering and guardrails are essential

02 **Sensitive Information Disclosure |**
Least-privilege for AI and continuous data-access monitoring

03 **Supply Chain |**
Third-party AI usage discovery and governance

04 **Data and Model Poisoning |**
Data lineage, training data controls, and approval workflows

05 **Improper Output Handling |**
Output sanitization and DLP-esque redaction on AI responses

06 **Excessive Agency |**
Agent mapping, runtime policy controls, and kill-switches

07 **System Prompt Leakage |**
Governance controls around data and secrets

08 **Vector and Embedding Weaknesses |**
AI-specific data lineage, classification, and governance.

09 **Misinformation |**
Human-in-the-loop, policy-based runtime controls

10 **Unbounded Consumption |**
Rate limits, budget caps, anomaly detection, and DoS-style defense

Read more at OWASP

Security Status

Once teams understand their AI inventory and the business purpose of each tool, they have the context to dive into the security status or posture. First, risk is mapped to the overall number of sensitive records associated with access to or from the AI tool. In addition, risk is also mapped to the OWASP Top 10 for LLMs and GenAI applications. From there, users can investigate specific AI assets and uncover more specific risk details such as:

Issues & Alerts: Security alerts, misconfigurations, or policy violations related to an AI asset.

Severity Score: Sort by low, medium, high, and critical issues/alerts.

Identity Mapping: Information about identity types, associated identity providers, trust level, and what data access they have within the AI.

Data Mapping: Drill down into the AI asset to investigate the associated datastores, data class, data type, and number of records at risk along with samples.

The security status experience for AI-SPM allows security teams to zoom in and out of their AI deployments with ease and better understand their environments. With AI-SPM they'll be able to test access controls, configurations, and policies to limit risk while enabling the business. It provides them with a holistic view to guide and prioritize their remediation efforts. Most importantly, AI-SPM will allow security teams to apply and enforce their organizational AI security policy.

Preventing Attacks with AI Runtime Protection

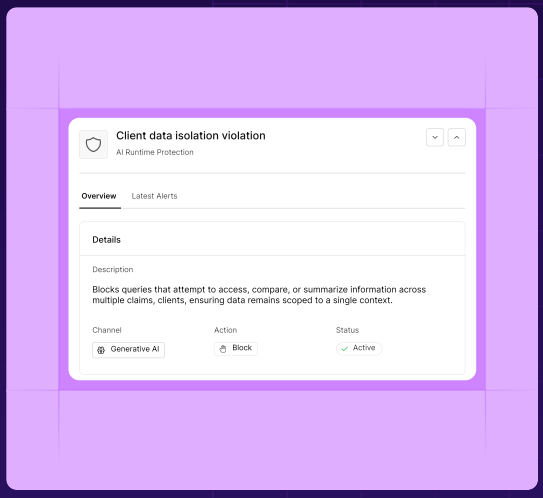
AI usage and agentic workflows are causing businesses to move faster than ever before—and security must keep up. That's where Cyera AI Runtime Protection comes in by enabling a safe roll out of AI at scale, while ensuring the right guardrails are in place in the event of a cybersecurity incident. The main capabilities included are **governance and enforcement.**

Governance

Before actively blocking threats, it's important to determine the guardrails for AI. This is done via policy controls leveraging Cyera's Omni DLP. These governance policies can be specific for AI such as blocking sensitive data in prompts, enforcing access permissions to models, preventing access to unapproved applications, and more.

Enforcement

Enforcement is a key capability missing in most AI security tools available today. Runtime Protection provides organizations with the ability to prevent AI attacks from occurring. Security teams can enforce policies on the spot to block or sanitize prompts and responses before they reach a specific model or user. It also correlates these runtime events to AI inventory to surface anomalies and risk patterns, which gradually improve the policies over time.



Security for AI Fundamentals

Want to get up to speed on AI Security? Sign up for Security for AI Fundamentals, a vendor agnostic educational course for securing AI in the enterprise. We go through the different aspects of AI, including new protocols, frameworks, and regulatory pressures. Whether you're a CISO, security engineer, system admin, or a non-technical stakeholder, this is an excellent way to keep pace with the industry.

Securing AI Is a Data Security Journey

Organizations across industries are investing heavily in AI, yet many are struggling to achieve meaningful results. Unlike previous technology shifts, AI places data, and particularly sensitive data, at the center of both opportunity and risk. To capture the promise of AI, data cannot be locked away in isolation. It must be secured while still remaining accessible to those who need it.

A successful AI strategy weaves security into every stage. Enterprises must identify data silos, classify the information fueling their AI, and apply guardrails that protect without slowing innovation. They also need a complete view of their AI deployments that grows and adapts as new data sources, models, and threats emerge.

> The leaders in AI will be those who treat data security as the foundation of innovation.

Cyera AI Guardian provides the foundation for this future. Built on proven strengths in data security, governance, and risk management, AI Guardian unifies AI Security Posture Management with AI Runtime Protection. It gives organizations visibility into every AI asset, real-time safeguards to prevent misuse, and the confidence to scale AI responsibly. With Cyera, enterprises can accelerate AI innovation while maintaining trust and control, ensuring that progress neve r comes at the cost of protection.