Guide

What Is AI Security Posture Management (AI-SPM)?

uptycs

Table of Contents

Key Takeaways

- AI-SPM helps organizations discover and assess risks across AI models, services, data, development pipelines, and supporting cloud infrastructure.
- It provides visibility into risks such as shadow AI, exposed services, excessive permissions, vulnerable components, sensitive data exposure, and AI supply-chain weaknesses.
- AI-SPM helps teams understand relationships and data flows across models, data sources, pipelines, identities, APIs, and infrastructure.
- Continuous discovery and contextual prioritization help organizations manage AI security risks as environments change.
- Governance, compliance, and clear ownership are important to maintaining a secure and trustworthy AI environment.
- AI-SPM works alongside CNAPP and other cloud security capabilities rather than replacing them.
- Effective AI security requires visibility across AI assets and the infrastructure that supports them.

AI Security Posture Management, or AI-SPM, helps organizations discover, assess, prioritize, and address security risks across AI models, data, services, development pipelines, and the cloud infrastructure supporting them.

As organizations adopt generative AI and machine learning, they introduce new assets, dependencies, and data flows that may not be fully visible through traditional cloud security processes. AI-SPM provides a continuous approach to understanding these environments, identifying weaknesses, and improving security throughout the AI lifecycle, from development and training through deployment and ongoing operation.

What Is AI-SPM?

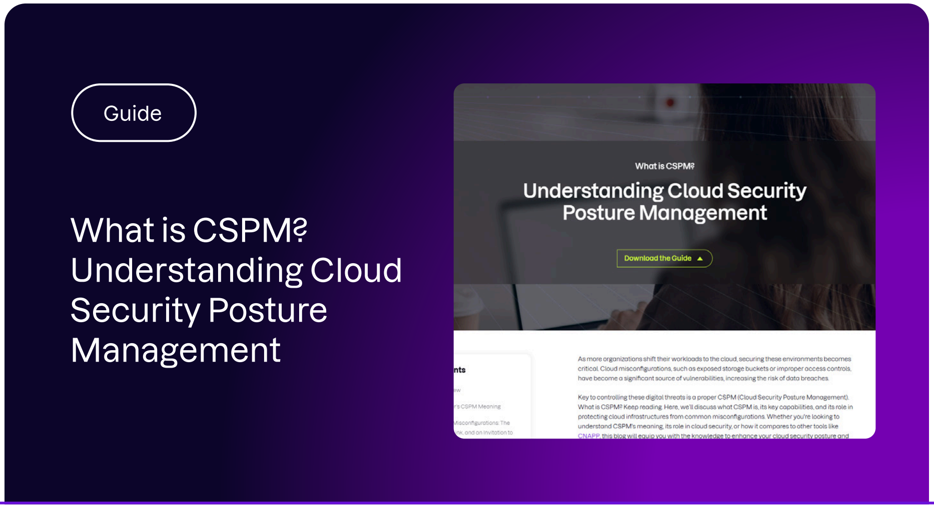
AI Security Posture Management is a continuous approach to identifying AI-related assets, evaluating their configurations and exposure, understanding how they interact with data and infrastructure, and prioritizing the risks that matter most.

The scope of AI-SPM extends beyond a model itself. Modern AI systems depend on data pipelines, APIs, identities, development tools, cloud services, containers, storage, networks, and other resources. A weakness in any of these connected components can affect the security, integrity, or availability of the AI application.

AI-SPM gives security and governance teams a clearer view across:

- Proprietary, open-source, third-party, foundation, and large language models
- Training, fine-tuning, grounding and inference data
- AI applications, APIs, agents, services, and model endpoints
- Notebooks, repositories, development pipelines, and model registries
- Packages, libraries, container images, datasets, and software dependencies
- Cloud infrastructure, identities, storage, containers, Kubernetes, serverless resources, and networks supporting AI workloads
- Relationships and data flows between models, data, services, identities, and infrastructure

AI-SPM supports security throughout the AI lifecycle, including development, training, deployment, and operation. It differs from controls focused only on testing model behavior or protecting the application interface. Its primary role is to provide continuous visibility into AI assets, posture, access, exposure, data usage, governance, lineage, and infrastructure context.



Why Is AI-SPM Important?

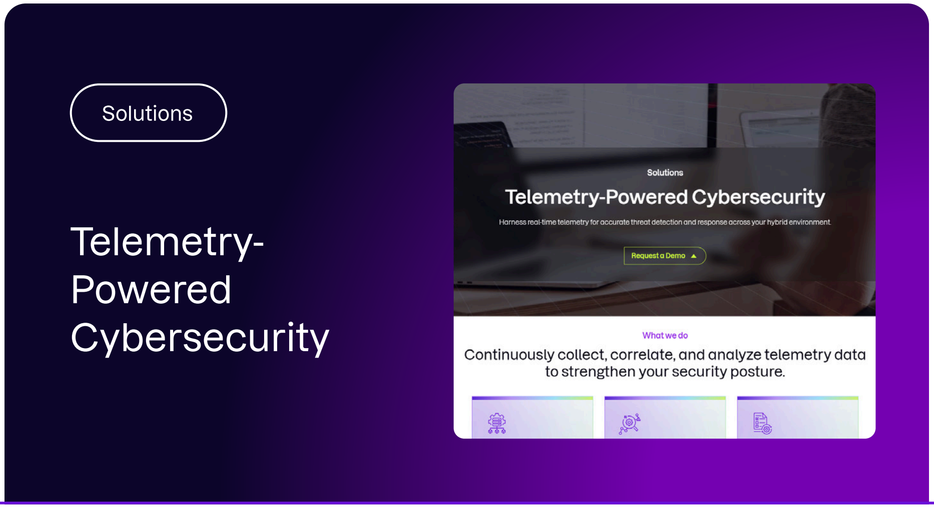
AI adoption is moving quickly across enterprises. Teams are experimenting with new models, services, agents, and data sources often using cloud resources that can be created or changed in minutes. This pace can leave security teams without a complete view of what has been deployed, who owns it, what data it uses, or how it connects to the rest of the environment.

At the same time, AI workloads are increasingly tied to sensitive data and critical infrastructure. Models may access confidential or regulated information during training, fine-tuning, grounding, or inference. AI services may also rely on broad permissions, public endpoints, third-party APIs, and rapidly changing software dependencies.

These conditions create several challenges:

- Security teams may not know where AI tools, models, agents, and services are deployed.
- Unapproved or unmanaged adoption can create shadow AI outside established security and governance processes.
- Exposed endpoints, insecure configurations, vulnerable dependencies, and excessive permissions can increase the likelihood of compromise.
- Third-party models, datasets, libraries, and APIs can introduce provenance and supply-chain risks.
- AI-specific threats, including data poisoning, adversarial attacks, model extraction, and model manipulation, may not be fully recognized by conventional cloud controls.
- Privacy, security, and AI governance requirements increase the need for policy enforcement, auditability, ownership, and accountability.

Traditional posture tools remain important, but they may not provide enough AI-specific context about models, data lineage, dependencies, and lifecycle risks. AI-SPM helps close that visibility gap while bringing AI assets into existing security, governance, and risk-management processes.



Common Security Risks Affecting AI Workloads

Shadow AI and Unknown Assets

AI resources can be deployed outside approved processes or without security-team visibility. Unknown models, agents, services, notebooks, or APIs may lack an accountable owner, approved business purpose, security review, or ongoing monitoring.

Misconfigurations and Exposed Services

AI endpoints, notebooks, storage, model registries, APIs, and supporting cloud resources may be publicly exposed or insecurely configured. Weak authentication, inadequate encryption, excessive logging of sensitive information, and improper authorization settings can increase risk.

Excessive Permissions and Unauthorized Access

Users, service accounts, workloads, applications, and AI agents may have broader access to data, models, or cloud resources than necessary. Excessive privileges can contribute to model misuse, unauthorized changes, data exposure, credential theft, and lateral movement.

Sensitive Data Exposure and Data Governance Gaps

Training data, prompts, outputs, embeddings, reference data, and logs may contain confidential, proprietary, personal, or regulated information. Without clear data classification and lineage, organizations may not know where sensitive data is used, stored, or shared across AI workflows.

Vulnerable Components and AI Supply-Chain Risks

AI workloads depend on models, packages, libraries, containers, datasets, APIs, and development tools. Vulnerable or unverified components can introduce risk before an AI application reaches production. Weak provenance, version tracking, and dependency controls can also make it difficult to establish an asset's integrity and history.

Data Poisoning and Model Manipulation

Attackers may alter training, fine-tuning, or reference data to influence model behavior, introduce malicious patterns, or reduce the reliability of outputs. Unauthorized model changes can produce similar effects and may be difficult to identify without lineage and integrity controls.

Adversarial Attacks and Model Extraction

Carefully manipulated inputs can cause models to produce incorrect or unsafe results. Attackers may also probe outputs to infer model behavior, reproduce proprietary functionality, or steal intellectual property.

Runtime and AI Application Threats

AI workloads can face the same runtime threats as other cloud applications, including workload compromise, credential theft, malware, malicious network activity, and unauthorized access. They can also face application-level risks such as prompt injection, insecure output handling, malicious model invocation, and abnormal model interactions.

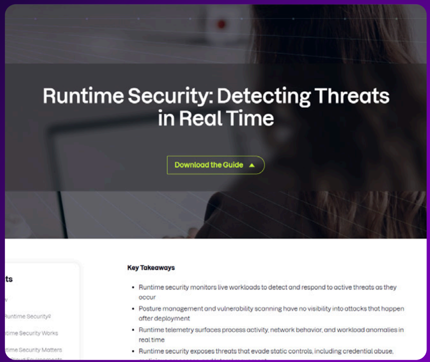
Compliance and Governance Risks

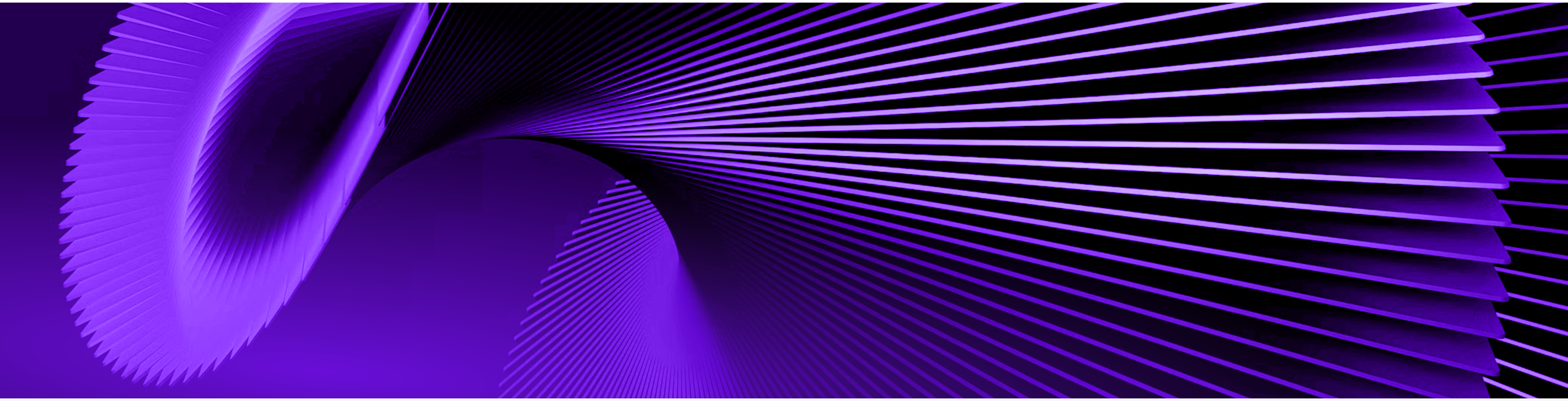
Organizations may lack approved-use policies, audit trails, ownership information, exception processes, or risk acceptance criteria for AI resources. These gaps make it harder to demonstrate compliance with privacy requirements, security standards, and emerging AI governance frameworks.

AI-SPM provides posture, exposure, access, governance, lineage, and infrastructure context for these risks. It does not replace runtime, application, data, or model security controls, which may also be required to detect and respond to active threats.

Guide

Runtime Security pillar page





How Does AI-SPM Work?

AI-SPM operates as a continuous cycle across AI assets from development through operation. Although individual platforms may organize capabilities differently, the process generally includes four stages.

1. Discover and Map

The first step is to identify AI-related assets and understand how they connect. This includes models, agents, endpoints, applications, APIs, data sources, vector databases, development tools, packages, container images, cloud services, identities, secrets, and runtime infrastructure.

Discovery should include approved and unapproved resources, as well as active, inactive, internally developed, open-source, and third-party assets. Each asset should be associated with an owner, business purpose, environment, and lifecycle status where possible.

AI-SPM also maps relationships between models, data sources, pipelines, APIs, identities, and cloud resources. This creates lineage and dependency context, such as model origin, versions, training data, deployment history, approvals, and ownership.

2. Assess

Once assets and relationships are understood, AI-SPM evaluates them for security and governance weaknesses. Assessments can include misconfigurations, public exposure, vulnerabilities, excessive permissions, insecure network access, sensitive data exposure, weak encryption or authentication, policy violations, and untrusted dependencies.

The assessment should consider both the AI-specific asset and the environment around it. A model endpoint may appear low risk in isolation, for example, but become more significant if it is internet-facing, connected to sensitive data, and accessible through an overprivileged identity.

3. Prioritize

AI environments can generate more findings than teams can address at once. AI-SPM prioritizes risks by correlating technical issues with context, including:

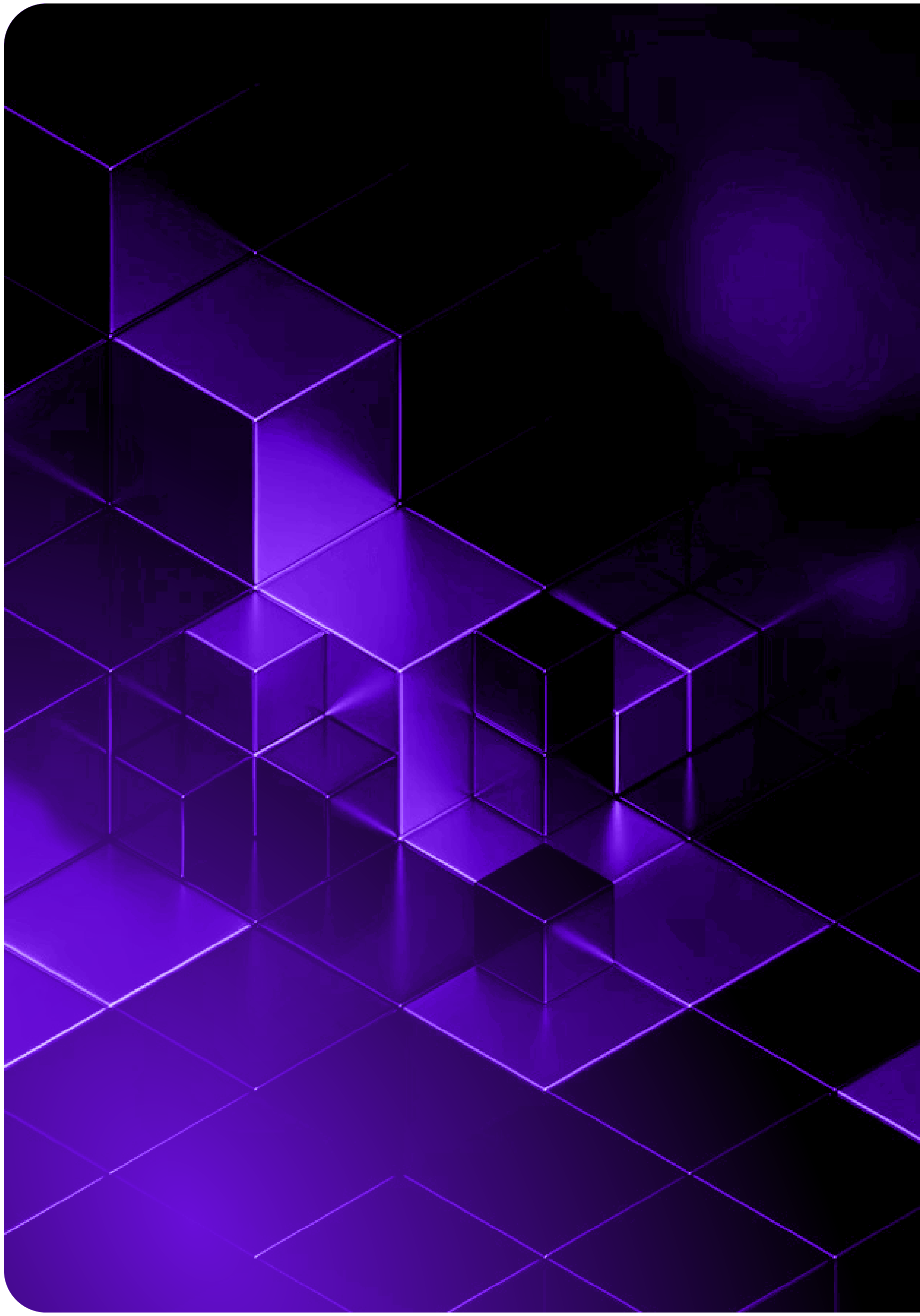
- Internet exposure and network reachability
- Asset importance and business purpose
- Identity and permission levels
- Data sensitivity and regulatory requirements
- Known vulnerabilities and exploitability
- Model lineage and supply-chain relationships
- Connections between AI services and cloud resources

- Potential attack paths
- Threat intelligence and available runtime signals
- Approval, ownership, and active-use status

This context helps teams focus on findings that are most likely to affect critical AI workloads, sensitive information, model integrity, or business operations.

4. Remediate, Govern, and Monitor
AI-SPM supports remediation by identifying the conditions that create risk and directing findings to the appropriate teams. Actions may include correcting insecure configurations, restricting public access, reducing permissions, addressing vulnerable dependencies, protecting exposed data, validating models and datasets, and assigning ownership to unmanaged assets.

Because AI environments change continuously, posture management cannot be a one-time assessment. Organizations need ongoing monitoring for new assets, changes in data flows, posture drift, emerging vulnerabilities, policy violations, and new relationships between models and infrastructure. Findings should integrate with development, MLSecOps, ticketing, governance, and incident-response workflows so teams can move from identification to resolution.



Best Practices for AI Security Posture Management

An effective AI-SPM program combines technology, governance, and repeatable security processes across the full AI lifecycle.

Maintain a Complete AI Asset Inventory

Continuously identify models, services, agents, datasets, development tools, endpoints, and supporting cloud resources. Record ownership, business purpose, approval status, environment, and lifecycle stage so unknown or unmanaged assets can be addressed.

Map AI Dependencies and Data Flows

Understand how models connect to datasets, APIs, identities, pipelines, applications, and infrastructure. Track model and data lineage to improve visibility, accountability, impact analysis, and incident response.

Protect Sensitive Data and Access

Classify sensitive data used by AI systems and apply controls for access, encryption, retention, logging, and movement. Enforce least privilege for users, workloads, service accounts, applications, and AI agents.

Secure AI Development and Supply Chains

Integrate security checks into development and deployment workflows. Validate third-party models, packages, datasets, containers, and other dependencies before use, and maintain provenance and version records.

Reduce Exposure and Misconfigurations

Review endpoints, notebooks, storage, registries, APIs, and supporting infrastructure for public exposure, insecure settings, weak authentication, and configuration drift.

Prioritize Risks Using Context

Use asset criticality, exposure, data sensitivity, permissions, vulnerabilities, attack paths, lineage, and runtime signals to determine which findings require the fastest response.

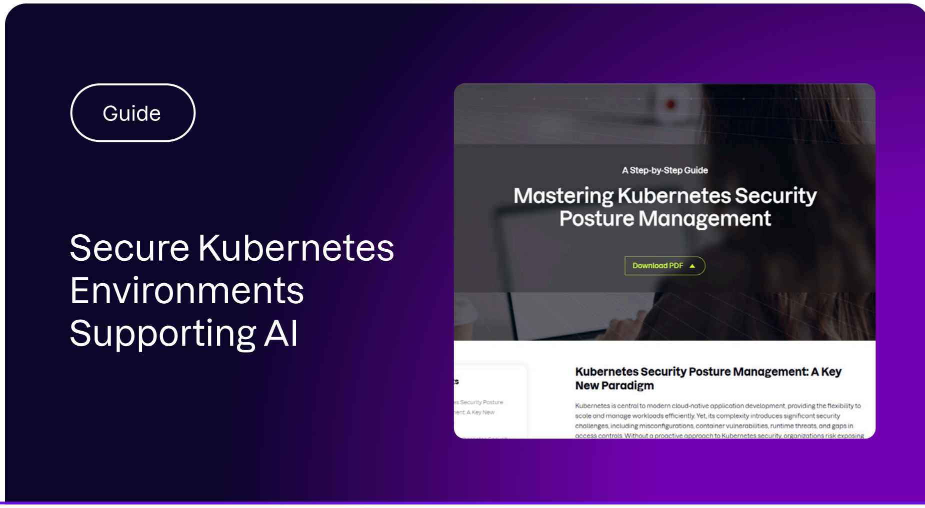
Establish Governance and Ownership

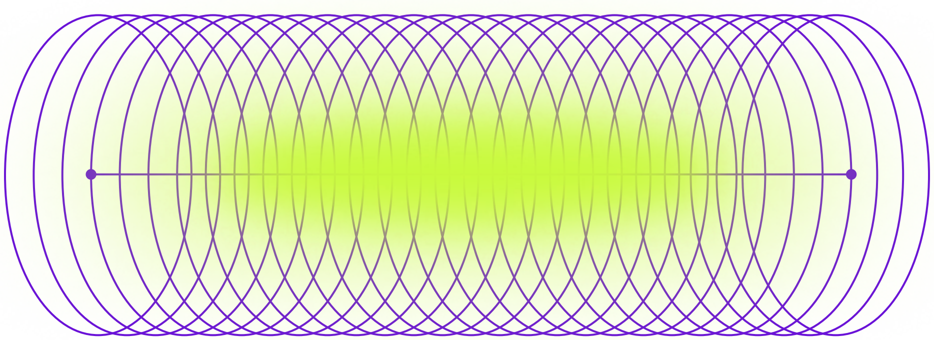
Define approved AI tools and practices, assign clear responsibilities, and maintain policies, exceptions, audit records, approval decisions, and risk acceptance criteria.

Continuously Monitor and Prepare for Incidents

Monitor for new assets, posture changes, sensitive data exposure, suspicious activity, and emerging vulnerabilities. Include AI assets, data flows, dependencies, and owners in incident-response planning.

AI-SPM should be part of a layered security strategy that also includes cloud, workload, application, data, identity, runtime, supply-chain, and model security controls.





How AI-SPM Fits Into a CNAPP Strategy

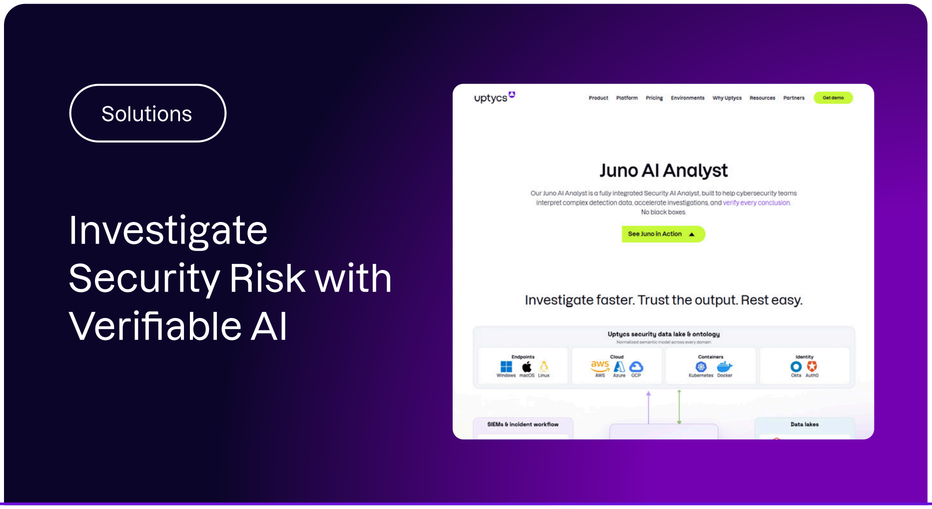
AI-SPM does not operate in isolation. AI workloads depend on the same cloud infrastructure, identities, storage, containers, Kubernetes environments, networks, APIs, and development pipelines protected by a cloud-native application protection platform, or CNAPP.

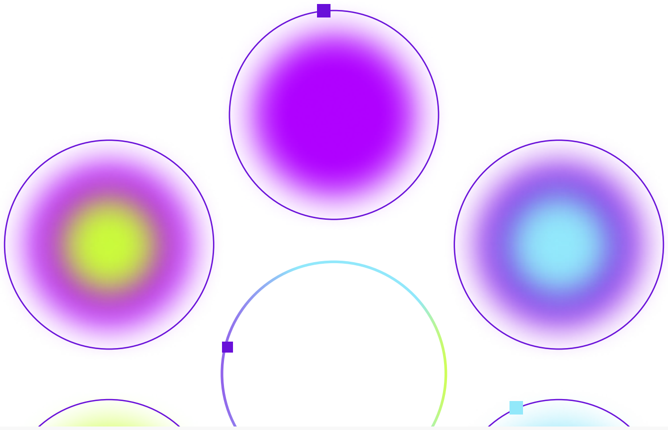
AI-SPM extends this visibility with AI-specific asset discovery, data lineage, model context, governance, and lifecycle risk. CNAPP provides broader protection across cloud posture, workload vulnerabilities, identities and permissions, containers, Kubernetes, network exposure, attack paths, development environments, and runtime threats.

Together, AI-SPM and CNAPP can help teams answer important questions:

- Which AI assets are deployed across the organization?
- What data, identities, applications, and infrastructure do they depend on?
- Who or what can access them?
- How do security issues connect across the environment?
- Which findings present the greatest risk?
- Is suspicious activity affecting an exposed or vulnerable AI workload?

AI-SPM complements CNAPP by extending posture and risk visibility to AI-specific assets, data flows, dependencies, and development lifecycles. It does not replace established cloud security controls. It also works alongside data security posture management, which focuses on sensitive data, and cloud security posture management, which focuses on the configuration and compliance of the underlying cloud environment.





How Uptycs Helps Protect AI Workloads

Securing AI involves more than protecting a model. AI applications depend on data, identities, APIs, development pipelines, cloud services, containers, Kubernetes environments, endpoints, and runtime infrastructure. Uptycs helps organizations protect this surrounding environment by bringing posture, identity, vulnerability, exposure, attack-path, and runtime context together in one platform.

Uptycs provides hybrid and multi-cloud visibility, cloud posture management, identity and entitlement analysis, vulnerability prioritization, container and Kubernetes security, and runtime workload protection. This context helps teams identify security issues in the infrastructure supporting AI applications and understand how configurations, identities, vulnerabilities, and exposure combine to create risk.

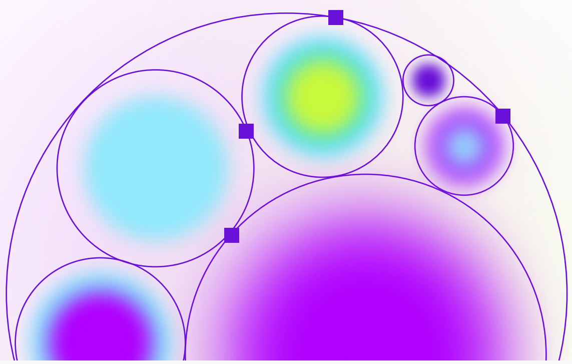
Investigate AI Workload Risk with Juno AI Analyst

Juno AI Analyst is integrated into the Uptycs platform to help security teams investigate detections and risks across cloud, Kubernetes, workload, endpoint, and runtime context. Rather than relying on an unsupported conclusion, analysts can review the queries, evidence, and reasoning behind each finding.

For environments supporting AI workloads, Juno can help teams investigate questions such as which assets are affected by a vulnerability, whether suspicious activity is connected to a cloud identity or workload, how a misconfiguration contributes to an attack path, and what evidence supports the recommended next step.

Juno can interpret complex detections, correlate signals across unified telemetry, identify what is affected and what is not, and present findings in clear language. This can reduce the manual effort required to move between posture findings, vulnerabilities, alerts, and runtime activity while keeping analysts in control of the investigation.

Every Juno output is designed to be verifiable, with the underlying evidence and query logic available for review. This is especially important when investigating AI-supporting infrastructure, where teams need to distinguish theoretical exposure from active risk and avoid making decisions based on a black-box answer.

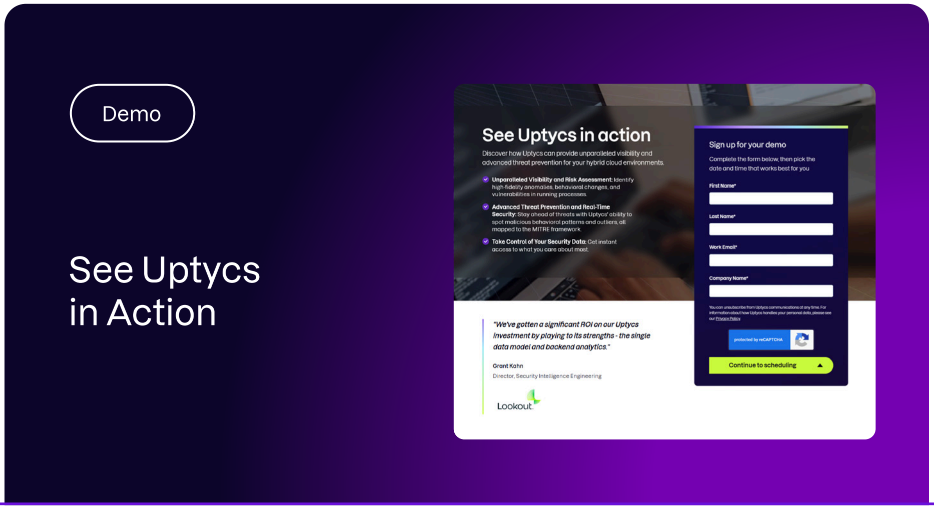


Connect Posture Findings with Runtime Context

AI-SPM helps organizations understand AI-specific assets, dependencies, data flows, and posture risks. Uptycs complements that view with broader cloud, identity, workload, and runtime context. By connecting preventive findings with active behavior, teams can better prioritize the conditions that need immediate attention and investigate suspicious activity affecting the infrastructure on which AI applications depend.

This combined context supports faster coordination across cloud, application, security, data, and operations teams. It also helps organizations build a stronger security foundation for AI without treating AI workloads as isolated from the rest of the environment.

Explore how Uptycs unifies cloud posture, identity, vulnerability, and runtime security across modern cloud environments.



Frequently Asked Questions

What does AI-SPM stand for?

AI-SPM stands for AI Security Posture Management. It is a continuous approach to discovering AI assets, evaluating their security posture, prioritizing risk, and monitoring changes across models, data, services, development workflows, and supporting infrastructure.

What is AI Security Posture Management?

AI Security Posture Management helps organizations identify and manage security, exposure, access, governance, and compliance risks across AI systems. It provides visibility into AI assets and the cloud, data, identity, and development resources they depend on.

Why do organizations need AI-SPM?

Organizations need AI-SPM because AI adoption can create new assets, data flows, dependencies, and risks faster than traditional inventory and review processes can track. AI-SPM helps reduce blind spots, identify shadow AI, and bring AI resources into established security and governance programs.

What types of assets does AI-SPM protect?

AI-SPM can cover models, agents, services, endpoints, APIs, datasets, vector databases, notebooks, pipelines, model registries, packages, containers, cloud services, identities, storage, and networks supporting AI workloads.

What is shadow AI?

Shadow AI refers to AI tools, models, services, or applications used without formal approval, oversight, or security visibility. These assets may process sensitive data or connect to business systems without appropriate controls.

What is AI lineage, and why is it important?

AI lineage records where a model and its data came from, how they changed, what dependencies they use, where they were deployed, and who approved or owns them. This context supports integrity checks, impact analysis, governance, and incident response.

How does AI-SPM support AI governance and compliance?

AI-SPM can help enforce policies, document ownership, identify unapproved resources, track data and model lineage, maintain audit records, and detect conditions that may conflict with security, privacy, or AI governance requirements.

How does AI-SPM protect sensitive data?

AI-SPM helps teams discover where sensitive data is used across training, fine-tuning, grounding, and inference workflows. It can identify exposure, excessive access, insecure storage, and data-flow risks so appropriate data security controls can be applied.

What is the role of AI-SPM in the AI development lifecycle?

AI-SPM provides continuous visibility from development and training through deployment and operation. It helps teams assess models, data, packages, pipelines, configurations, permissions, and infrastructure before and after an AI application reaches production.

How is AI-SPM different from CSPM?

CSPM focuses on the configuration, exposure, and compliance of cloud infrastructure. AI-SPM adds AI-specific visibility across models, data, agents, services, pipelines, lineage, and lifecycle risks. The two capabilities are complementary.

How is AI-SPM different from DSPM?

DSPM focuses on discovering, classifying, and protecting sensitive data. AI-SPM focuses on the broader posture of AI systems, including models, data usage, permissions, services, dependencies, and supporting infrastructure. AI-SPM can use data sensitivity context from DSPM to improve risk prioritization.

How does AI-SPM fit into a CNAPP strategy?

AI-SPM extends CNAPP visibility to AI-specific assets and relationships. CNAPP secures the broader cloud environment, including posture, workloads, identities, vulnerabilities, containers, attack paths, and runtime activity.

Can AI-SPM help protect generative AI workloads?

Yes. AI-SPM can help discover generative AI models, services, agents, data sources, vector databases, endpoints, and supporting resources. It can then assess posture, exposure, access, lineage, governance, and infrastructure risks across those environments.

Can AI-SPM detect data poisoning and model extraction risks?

AI-SPM can identify conditions associated with these risks, such as untrusted data sources, weak lineage, excessive access, exposed endpoints, and missing integrity controls. Specialized data, application, model, and runtime security capabilities may still be required to detect active attacks.

Does AI-SPM protect AI models from every type of attack?

No. AI-SPM primarily addresses discovery, posture, exposure, access, data usage, governance, lineage, and infrastructure risk. Organizations should combine it with runtime, application, data, identity, supply-chain, and model security controls to address threats such as prompt injection, adversarial attacks, and model manipulation.

Uptycs delivers the industry's first **Unified CNAPP** with integrated XDR, powered by Juno - the only verifiable AI Security Analyst.

Unlike traditional CNAPPs that rely on static snapshots, Uptycs unifies cloud configuration data with deep runtime telemetry across cloud, containers, and endpoints into a single Unified Ontology. Trusted by the Fortune 500, Uptycs powers Continuous Threat Exposure Management (CTEM) by normalizing telemetry into a single data lake, enabling the Juno AI Analyst to provide **"Glass Box" transparency**, turning hours of investigation into minutes of verified answers.

Learn more at uptycs.com/juno-ai