

WHITEPAPER

Securing Enterprise AI at the Speed of Adoption

Bring AI activity into the SOC—see risk earlier, make defensible decisions, and act with control.

securonix

Table of Content

Executive Summary	3
AI Changes The Trusted-Access Problem.....	4
Find Shadow AI Before It Becomes a Data or Access Incident.....	5
Know When a Trusted Enterprise Agent Moves Outside its Expected Purpose	7
Detect the Administrative Changes That Reshape AI Risk.....	8
Connect Weak Signals Before AI Risk Becomes an Incident	9
Use Existing Telemetry to Close AI Visibility Gaps.....	10
Measure Decision Readiness Not Alert Volume.....	13
Make AI Activity Part of the SOC Security Story.....	14

Executive Summary

Bring AI activity into the SOC before visibility and control fall behind adoption.

AI adoption is moving faster than many security teams can inventory services, validate controls, or connect activity to accountable owners. Yet the core decision remains familiar: what changed, who or what caused it, which data and systems were affected, and what response is justified?

Copilots, model platforms, retrieval-augmented generation, and autonomous or semi-autonomous agents expand the attack surface beyond prompts and models. Risk can originate in cloud control planes, human and non-human identities, application consent, sensitive repositories, endpoints, APIs, and network routes. An attacker, over-permissioned user, or agent can create material exposure without compromising the model itself.

Security leaders need one connected operating model that brings AI activity into the evidence, investigation, response, and governance workflows their SOC already uses. This approach helps analysts detect risky change sooner, reach defensible decisions faster, and preserve the cases, data controls, and approval structures that keep response accountable as adoption changes.

43

out-of-the-box AI threat policies at publication

6

coverage domains across platform, identity, data, endpoint, and network

2 WEEK

target cadence for expanding AI threat coverage

AI Changes The Trusted-Access Problem

Approved users and authorized agents can still create risk. The SOC must judge behavior in context—not simply allow or block access.

Enterprise AI expands the number of trusted actors and the speed at which they can retrieve, transform, and move information. A user may upload a regulated file to a public assistant. An administrator may weaken a cloud AI guardrail. An application may receive high-scope delegated permissions. An AI agent may be granted a sensitive role, call an unapproved tool, or act from an unusual location. Each event may be legitimate in isolation. That ambiguity is why model-only controls and static allowlists are insufficient. Security teams need evidence from the systems around the model, plus a baseline for what is normal for the person, service principal, agent, application, device, resource, and business process.

RISK PATTERN	EVIDENCE TO CONNECT	DECISION TO MAKE	SECURONIX RESPONSE
Platform integrity	Cloud audit, model, guardrail, logging, agent, and knowledge-base changes	Can the organization prove the control-plane change was expected and approved?	Detect configuration drift and connect the change to the identity, affected resource, approval context, and prior activity so analysts can determine whether it was expected and approved.
Identity & permission abuse	Sign-in, consent, role assignment, service principal, device posture	Does the identity or agent need this level of access now?	Baseline human and non-human entities and escalate rare or high-scope changes that require review.
Anomalous AI use	Interaction, invocation, location, volume, failure, and model-use metadata	Is this legitimate adoption or account misuse?	Compare activity with peer, historical, and environmental behavior to distinguish adoption from misuse.
Data exposure	DLP, file, repository, sensitivity, upload, destination, transfer volume	What data moved, to where, and under whose authority?	Correlate data and identity evidence so analysts can prioritize movement that violates policy.
Endpoint execution	Process, shell, DNS, script, connection, parent-child process	Is an AI API part of an execution chain?	Connect endpoint activity to AI destinations and the broader execution chain.
Network transfer	Proxy, firewall, destination rarity, failed access, bytes sent	Is this sanctioned use, compromised access, or exfiltration?	Surface unusual destinations and outbound volume, then add user and host context for a defensible decision.

Find Shadow AI Before It Becomes a Data or Access Incident

See unsanctioned AI use sooner. Focus investigation on policy-breaking access or data movement—not every new AI connection.

THREAT

AI Adoption Outruns Inventory

Employees and developers can reach public AI services, browser extensions, APIs, and local agents before formal review or logging is in place.

EVIDENCE

The Evidence is Fragmented

Proxy, firewall, endpoint, DNS, identity, SaaS audit, and DLP evidence must be connected to show who connected, from where, and what moved.

DECISION

Rarity Does Not Reveal Intent

A rare AI destination may reflect research, pilot, a compromised endpoint, or policy-breaking transfer. Analysts need context to distinguish them.

Help analysts distinguish experimentation from exposure

Analysts need to separate legitimate experimentation from policy violations, compromised access, and unauthorized data movement without pivoting across disconnected tools. The Threat Detection for AI Value Package includes policies for rare AI destinations, repeated failed connections, files uploaded to AI sites, abnormal transfer volume, and endpoint processes connecting to AI APIs. Unified Defense SIEM normalizes the supporting telemetry, while UEBA establishes user and host baselines. Analysts can then correlate the destination with sign-in history, device posture, file sensitivity, transfer volume, and other policy violations.

- Prioritize systems and identities that policy prohibits from using public AI services.
- Distinguish a first-time business user from a scripted process or compromised host.
- Move confirmed violations into the SOC's established case and response workflow.
- Use Data Pipeline Manager and the Data Pipeline Agent to preserve relevant visibility while aligning retention and processing with analytic value.

The result is earlier discovery, fewer manual pivots, and investigations centered on policy-breaking access or data movement rather than every new AI connection.

Stop AI Assisted Work From Crossing Data Boundaries

Determine what data moved, who or what moved it, and whether the transfer violated policy. Then respond without losing accountability.

AI accelerates the movement of sensitive information. Users can upload source code, customer records, financial information, or internal documents to assistants. Agents can retrieve content from knowledge bases at machine speed. Retrieval-augmented generation workflows may expose data through over-broad retrieval, compromised identities, or manipulated configuration. Security teams must determine what data was involved, who or what moved it, where it went, and whether that movement complied with policy.

RISK PATTERN	EVIDENCE TO CONNECT	DECISION TO MAKE	SECURONIX RESPONSE
Risky file upload	SharePoint audit, DLP alert, user, file type, label, destination	Was the upload permitted for this data class?	Correlate file and identity context. Raise priority when sensitive data conflicts with the destination or applicable policy.
Abnormal transfer	Proxy/firewall bytes, destination, device, prior usage, time	Does the volume fit this user, host, and workflow?	Use behavioral analytics to identify unusual outbound volume to AI services and focus analyst attention.
RAG extraction	Knowledge-base access, model invocation, identity, request pattern	Is retrieval consistent with role and intended application behavior?	Detect potential extraction patterns and connect them to identity, access, and control-plane changes.
Agent overreach	Tool calls, delegated permissions, repository access, case history	Did the agent exceed its approved purpose?	Use AEBA to place agent actions in behavioral and enterprise context so the SOC can assess whether the agent exceeded its purpose.

Contain exposure without giving up control

Policy-based containment must protect privacy, approvals, and ownership. Sam can assemble the investigation and recommend next steps. Response Agent and SOAR can invoke approved playbooks—such as disabling a token, containing a host, revoking consent, or opening a data-loss workflow, while preserving an auditable record. Analysts remain accountable for consequential decisions.

Know When a Trusted Enterprise Agent Moves Outside its Expected Purpose

Treat every enterprise agent as a non-human identity with a defined purpose, bounded access, observable behavior, and an accountable owner.

AI agents can retrieve documents, invoke tools, write to systems, and act through delegated permissions. Their speed and access can turn excessive authority or behavioral drift into material exposure before a human notices. Risk may begin with an agent created by a rare administrator, high-scope application consent, a sensitive cloud role, long-lived credentials, an unusual sign-in, or tool use outside the agent's intended purpose. Security teams therefore need to govern agents as non-human identities.

KNOW

Assign accountable ownership

Record the business owner, purpose, approved tools, data boundaries, identities, environments, and response contact for each production agent.

BASELINE

Detect drift from intended behavior

AEBA extends the UEBA behavioral foundation to enterprise AI agents, helping teams assess expected purpose, own or inherited authority, tools, data access, interactions, and behavioral drift.

CONTROL

Keep authority bounded

Agentic Guardrails apply role-based access, approval gates, separation of duties, action limits, transparency, and audit trails to Securonix agentic workflows.

Apply the same identity and behavior discipline to agents that the SOC uses for people and service accounts. The outcome is faster agent-risk detection and greater SOC capacity without weakening human control.

Detect the Administrative Changes That Reshape AI Risk

See who changed logging, guardrails, models, agents, and connected data before visibility or trust degrades.

Administrative changes can silently reshape AI exposure. Logging can be reduced. Guardrails can be changed. Models can be deployed or fine-tuned. Data sources and knowledge bases can be attached. Agents can be created or updated. Because these actions are often legitimate, security teams need context rather than a binary alert.

Representative Coverage

1 Protect visibility into AI administration

Surface changes that reduce visibility or alter audit configuration. Validate the actor, change window, affected resource, and resulting evidence gap.

2 Validate changes that alter AI behavior

Detect modifications to Bedrock guardrails, model deployment, fine-tuning, or resource configuration and correlate them with privileged identity activity.

3 See when connected data increases exposure

Identify new or modified associations and potential RAG-based extraction behavior. Connect resource changes to data access and invocation history.

4 Connect agent changes to ownership and behavior

Detect new or modified agents, tools, roles, and associated permissions. Connect the change to the administrator, approved purpose, and subsequent agent activity.

Why context changes the decision

A guardrail update during an approved release carries a different risk than the same change made by a new administrator at an unusual time and followed by high-volume retrieval. Securonix correlates the control-plane event with identity, peer behavior, location, device, and follow-on activity so analysts can judge the full sequence rather than a single event.

Connect Weak Signals Before AI Risk Becomes an Incident

See behavior that looks ordinary in one tool but becomes material across an identity, endpoint, data, and network timeline.

Material AI risk often begins with a weak behavioral signal: a first-time user generating unusually high volume, an approved application accessed from a rare country, repeated model-invocation failures, a script connecting to an AI API, or a local agent spawning a shell. No single event settles the question. Analysts need connected context and careful tuning to avoid missing misuse or escalating legitimate work.

ANOMALOUS USAGE

Prioritize meaningful change

Baseline first-time use, rare identities, unusual locations, failed operations, model rarity, and interaction volume. Tune for planned rollouts and seasonal work.

ENDPOINT EXECUTION

Reconstruct the execution chain

Connect process ancestry, script execution, shell behavior, DNS, and AI API connections. Separate ordinary browser use from programmatic activity.

NETWORK TRANSFER

Close gaps in native visibility

Use proxy and firewall evidence for unsanctioned destinations, repeated failure, and abnormal outbound volume when application-native visibility is incomplete.

Example Investigation Narrative

1

Rare Sign-In

A developer identity authenticates from an unusual geography.

2

Permission Change

The account grants high-scope consent to an AI-connected application.

3

Endpoint Activity

A script resolves and connects to an AI API.

4

Data Movement

Proxy telemetry shows abnormal outbound volume.

5

Decision

Sam assembles the evidence into a case; the analyst decides whether a response is justified and approves the governed action.

Each event is ambiguous alone. Together, the sequence gives analysts enough context to distinguish misuse from legitimate adoption, focus the investigation, and choose a proportionate response.

Use Existing Telemetry to Close AI Visibility Gaps

The Securonix Threat Detection for AI Value Package adds detection content to the SOC's current data pipeline rather than creating another AI-only monitoring silo.



Detection is credible only when the source data is ready

A policy cannot detect an event that the source platform never produced or the SIEM never parsed. Before claiming coverage, confirm that each required connector is active, relevant logging is enabled, parsing is stable, and the fields used by the policy are populated. Some sources require additional configuration, including AWS Bedrock data-event selectors, Microsoft audit and diagnostic settings, and Google Cloud Data Access logs.

RISK PATTERN	EVIDENCE TO CONNECT	DECISION TO MAKE	SECURONIX RESPONSE
AWS Bedrock	CloudTrail, including required data-event selectors	Guardrail, logging, agent, model, and knowledge-base changes	Enable the required source events and validate coverage against a known administrative action.
Microsoft AI	Entra ID, sign-in, M365 audit, SharePoint, Azure Monitor	Consent, permission, Copilot use, file upload, Foundry activity	Confirm workload licensing, audit configuration, and diagnostic settings
Google Cloud AI	GCP Audit and Data Access logs	Agent activity, IAM, sensitive roles, data access	Verify Data Access logging where needed and scope the relevant projects and resources.
Endpoint / Network	Sysmon, proxy, NGFW, DLP	Scripts, local agents, AI destinations, transfers	Check parser compatibility and verify actual event-field coverage.

This readiness work prevents a missing log or empty field from being mistaken for a clean security result.

Turn AI Alerts into Decision-Ready Cases

Accelerate investigation and containment while preserving privacy, approval, reversibility, and ownership.

AI alerts create work; they do not settle the decision. Securonix connects detection and response inside the operating model the SOC already uses. Sam can enrich and prioritize the alert, build a timeline, summarize the evidence, and prepare a case for escalation. Threat intelligence adds external context. Analysts can examine identity, agent, asset, data, endpoint, cloud, and network evidence in one workflow, then decide whether action is justified.

Faster Decisions With Human Control

WORKFLOW STAGE	EVIDENCE TO CONNECT	DECISION TO MAKE	SECURONIX RESPONSE
Triage	Policy violation, severity, behavior, peer context	Does this deserve investigation now?	Sam prioritizes the alert and explains the evidence used so analysts can focus on material risk.
Enrichment	Identity, device, data, asset, threat intelligence	What changed, what is affected, and what remains unknown?	Sam and specialized agents gather and organize relevant context into a decision-ready case.
Decision	Risk narrative, response options, business ownership	Which action is proportionate and permitted?	The analyst validates the recommendation against policy, business ownership, and approval rules.
Execution	Approved playbook, action scope, rollback path	Can containment proceed automatically or require approval?	Response Agent and SOAR execute only the steps permitted by policy and approval boundaries.
Audit	Evidence, recommendation, approver, action, result	Can the organization defend the decision later?	The workflow retains the evidence, recommendation, approval, action, and result needed to defend the decision.

Apply policy, privacy, approval, and ownership rules before responding to AI-related threats. Response Agent and SOAR can invoke approved playbooks while analysts retain control of consequential decisions and the workflow preserves an auditable record.

A 90-Day Plan to Operationalize Governed AI Security

Prove the data, decisions, and controls before expanding automation.

DAYS 0-30

Expose gaps and assign ownership

Identify AI services and agents in production. Assign owners. Map policy requirements to available logs. Enable missing audit settings. Test known actions end to end.

DAYS 31-60

Establish normal behavior and tune coverage

Activate control-plane and permission policies. Establish behavior baselines. Document expected regions, administrators, agents, apps, data classes, and transfer patterns.

DAYS 61-90

Connect response and enforce governance

Add data, endpoint, proxy, and firewall coverage. Integrate cases with Sam and SOAR. Define approval gates, response boundaries, rollback, privacy review, and audit retention.

Operational disciplines

1 Validate telemetry before measuring detection

Confirm logs are enabled, collected, parsed, field-complete, and stable before using detection results to make coverage claims.

2 Match each policy to available data

Enable only policies whose source requirements can be met. Treat missing data as a coverage gap, not a benign result.

3 Baseline behavior before escalating anomalies

Account for rollouts, new business units, approved pilots, rare administrators, and changing model usage before escalating anomalies.

4 Bring evidence into one SOC workflow

Bring AI-related findings into established case management so analysts can correlate identity, data, cloud, endpoint, and network evidence in one investigation.

4 Govern automation before action

Define ownership, privacy controls, approval boundaries, separation of duties, reversible actions, and audit requirements before automated actions are enabled.

The result is a repeatable SOC capability that can absorb new AI services, agents, and telemetry without creating another set of disconnected controls.

Measure Decision Readiness Not Alert Volume

Show leaders whether the SOC has the visibility, speed, capacity, governance, and cost control to secure AI adoption.

OPERATING MEASURE	WHAT TO MEASURE	WHY IT MATTERS	EVIDENCE SOURCE
Visibility and coverage	Percent of in-scope AI services with validated logging and required fields	Shows where detection content can operate and where evidence gaps remain	Service inventory, connector health, parser and field validation
Detection	Time from meaningful change to correlated risk signal	Measures how quickly the SOC sees material AI-related activity	Policy timestamps, behavior events, correlation timeline
Decision	Time from signal to decision-ready case	Shows whether investigation workflows reduce friction and produce sufficient evidence	Sam investigation lifecycle, analyst touch time, evidence completeness
Response	Time from approval to completed and verified action	Tests execution speed, ownership, verification, and rollback readiness	SOAR and Response Agent action records
Quality	Escalation precision, reopen rate, and expected-change rate	Shows whether tuning distinguishes material risk from expected adoption	Case disposition, exception records, policy tuning
Governance	Percent of automated actions with required approval and audit evidence	Shows whether automation stays within approval and audit requirements	Guardrail, approver, action, result, and audit logs
Economics	Analyst effort per decision and telemetry cost per protected service	Connects AI defense to analyst capacity and the cost of maintaining useful visibility	Work effort, data routing, retention, processing, and service inventory

Board-ready questions



- Which AI services and enterprise agents are in scope, and which remain unknown?
- Can the SOC prove that high-impact control-plane, permission, and data-movement events are logged and detectable?
- How quickly can the team move from meaningful AI-related change to a defensible decision?
- Which response actions are automated, which require human approval, and how is accountability preserved?
- Does AI adoption force cost-driven visibility gaps, or is telemetry aligned to security value?

Make AI Activity Part of the SOC Security

Story

Secure AI adoption with a connected SOC that sees risk, explains decisions, and acts with control.

AI adoption can outrun the SOC's ability to see risk, connect evidence, and enforce accountability. Yet material threats still leave evidence across identities and permissions, cloud control planes, data repositories, endpoints, applications, and network paths. Isolating AI monitoring from those systems strips analysts of the context needed to make sound decisions.

Securonix brings AI-related activity into Unified Defense SIEM so analysts can investigate it alongside the identity, data, endpoint, cloud, application, and network evidence they already use. Out-of-the-box policies and behavior analytics surface material change. Threat-chain analytics and Sam assemble the evidence into a connected case. Response Agent and SOAR execute approved playbooks within human-defined boundaries. Data Pipeline Manager and the Data Pipeline Agent help preserve relevant telemetry while aligning processing and retention to analytic value.

The result is an operating model that can adapt as AI services change while improving visibility, decision speed, analyst capacity, accountability, and cost control.

Start With Evidence You Can Test

Inventory the AI services and agents already in production. Select the relevant Threat Detection for AI policies. Then validate one known administrative or user action for every required source. This exercise shows where coverage is credible, where evidence is missing, and which gap the team should close first. Learn more at securonix.com.