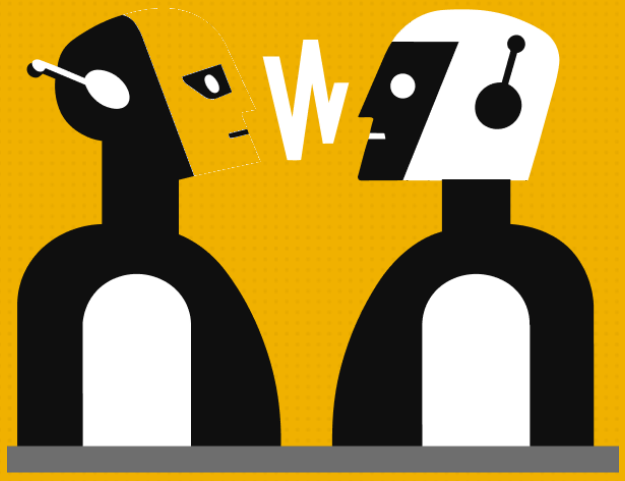
SOLUTION BRIEF

wallarm



AI Governance That Works at Runtime

How Wallarm Closes the Gap From Discovery Through Runtime Enforcement to Continuous Evidence

THE CHALLENGE

AI is core infrastructure. Nobody is watching it.

AI is now core infrastructure. Agents make decisions, access data, call external APIs, and act on behalf of users, often without anyone watching. Boards want assurance. Auditors want evidence. Security teams want control. And most organizations have none of the three.

The tools that exist today weren't built for this. Governance platforms catalog AI use cases and track policy. They don't see what AI is actually doing at runtime. Security tools detect threats after the fact. Neither can answer the questions that matter.

Q.01	What AI is running in our infrastructure right now, including the services nobody approved?
Q.02	Is sensitive data crossing boundaries it shouldn't?
Q.03	When an AI agent takes an action that violates policy, can we stop it in real time?
Q.04	If an auditor asks for evidence that our AI stack is governed, can we produce it today?

Governance without enforcement is documentation. Enforcement without visibility is guesswork. Most organizations have neither.

THE GOVERNANCE GAP

Building fast. Governing slow.

Sources: LangChain State of AI Agents; Gartner Cybersecurity Innovations in AI Risk Management Survey.

78% of organizations are actively building AI agents.

49% rely on manual reviews to detect embedded AI features.

No runtime enforcement. Most AI governance platforms describe what should happen. Few can stop it when it doesn't.

THE COST OF INACTION

What happens if you wait.

REGULATORY	EU AI Act violations carry fines of up to 3% of global annual revenue. SOC 2 and sector audits increasingly scope AI workloads. Evidence assembled after the fact rarely satisfies examiners.
BRAND	An agent that makes an unintended decision at scale (a jailbroken model, an over-permissioned role, a data boundary crossed) becomes a press incident before it becomes an internal ticket.
OPERATIONAL	Containment without runtime enforcement means pod restarts, credential rotation, and downtime. Every incident you respond to after the fact is one you failed to stop in real time.

WHAT THIS LOOKS LIKE IN PRACTICE

Five scenarios. Two outcomes.

WITHOUT WALLARM	WITH WALLARM AI GOVERNANCE
Security teams field alerts on AI workloads but can't explain what the AI was doing when the alert fired.	Every alert arrives with full behavioral context: which agent, which user, which call, what data was accessed, both before and after the event.
Shadow AI services spin up in Kubernetes clusters. Security finds out weeks later during an audit.	Wallarm auto-discovers every AI workload from runtime behavior. New services surface within minutes, including those that bypassed approval workflows.
Sensitive data flows through AI models and pipelines with no visibility into what crosses what boundary.	PII, credentials, and sensitive identifiers are detected in real time as they move through AI pipelines. Violations are flagged and blocked before they reach external destinations.
A jailbroken agent takes an unintended action. Containment means pod restarts, credential rotation, and an incident report explaining the gap.	Compromised sessions are revoked at the connection level in real time. No restart, no downtime. The agent stops. The record is complete.
Compliance evidence for SOC 2 or regulatory audits is assembled manually from logs and spreadsheets before the review.	AI inventory, session audit logs, coverage heatmap, and sensitive data flow records are generated continuously at runtime. Audit-ready at any moment.

WHY RUNTIME IS THE GAP

Policy without runtime is paperwork.

Traditional governance platforms (those covering inventory, policy templates, and risk assessment workflows) address the planning and documentation side of AI governance. That work matters. But Gartner's own market analysis of AI governance platforms identifies a consistent weakness across the category: most lack strength in runtime, in-line policy enforcement and protection.

That's the gap Wallarm was built to close. AI Hypervisor operates at the kernel level using eBPF and non-invasive memory analysis. It sees every outbound connection an AI workload makes, every piece of sensitive data that moves through an AI pipeline, and every policy violation, in real time, without touching application code. This is not a monitoring tool. It's a governance layer with teeth.

THE WALLARM APPROACH

The Wallarm AI Control Loop

Discover. Observe. Enforce. Govern.

Detection without enforcement is just a longer incident report. Wallarm makes it stop.

01 / DISCOVER

Discover

Know what AI is actually running.

AI Hypervisor auto-discovers every AI workload on your Kubernetes clusters, including shadow AI that bypassed approval. Infrastructure Discovery maps your full estate across accounts, regions, and environments. Within minutes you have a live topology.

02 / OBSERVE

Observe

See what AI is doing at runtime.

Captures every outbound connection (LLM calls, database access, external APIs, S3 reads) and attributes each back to the user or session that triggered it. Sensitive data is flagged in real time as it moves through pipelines.

03 / ENFORCE

Enforce

Stop violations before they become incidents.

When an AI agent violates policy, Wallarm blocks the outbound call at the connection level. Compromised sessions are revoked by user identity or trace ID. No pod restart. No deploy cycle. No downtime.

04 / GOVERN

Govern

Continuous, audit-ready evidence.

Generates a live AI inventory, a coverage heatmap scored against 63 controls, session audit logs, and sensitive data flow records, continuously, not the week before an audit.

CAPABILITY MAPPING

Requirements, mapped to Wallarm.

Effective AI governance requires a set of core capabilities. Here's how Wallarm delivers each, and which component carries it.

GOVERNANCE REQUIREMENT	HOW WALLARM DELIVERS IT	COMPONENT
AI Inventory and Catalog	Auto-discovers every AI workload from runtime behavior, including shadow AI, and maintains a live topology map of services, connections, and model calls.	AI HYPERVISOR INFRASTRUCTURE DISCOVERY
Risk Assessment & Scoring	Coverage heatmap scored across 63 controls, aligned to NIST AI RMF and SOC 2 control families. Updated continuously.	AI HYPERVISOR
Automated Runtime Enforcement	Blocks outbound LLM calls on pattern-match rules. Revokes compromised sessions at the kernel level by user or trace ID. Satisfies access control and session management requirements (NIST AC-2, SI-4). No application code changes required.	AI HYPERVISOR
Sensitive Data Governance	Detects PII, credentials, and sensitive identifiers in real-time AI traffic before they reach external destinations: credit cards, SSNs, passports, email addresses, API keys, JWT tokens. Addresses data boundary monitoring requirements across SOC 2 and NIST AI RMF.	AI HYPERVISOR
Evidence & Audit Trail	Generates AI inventory, session audit logs, sensitive data flow records, and coverage heatmap continuously. Reports are built at runtime, not assembled retroactively. Directly addresses documentation and logging obligations for high-risk AI systems.	AI HYPERVISOR
Supply Chain Visibility	A complete inventory of components and CVEs for every AI workload. Identifies vulnerable framework versions and high-severity CVEs in the live environment.	AI HYPERVISOR
Shadow AI Detection	Discovers unsanctioned AI services from behavioral signals (agent frameworks, MCP servers, vector databases, model provider calls) without relying on declared manifests. On detection: workload is logged to the topology graph, an alert fires, and OPA/Rego enforcement policies can block first-run instances of unapproved frameworks before they process any traffic.	AI HYPERVISOR
Interoperability	Integrates with SIEM, Security Hub, GuardDuty, and GRC platforms. Findings placed on the relationship graph with full behavioral context.	INFRASTRUCTURE DISCOVERY

WHO THIS IS FOR

Four roles. One platform.

CISO

AI risk exposure, agent behavior, compliance mandate.

- Context on every AI finding, not just alerts.
- Real-time blocking of PII egress and policy violations.
- Continuous evidence for board-level reporting.

CTO / VP Engineering

Deploy AI fast without accumulating security debt.

- Zero code changes; 60-second instrumentation.
- Governance that doesn't slow release velocity.
- Full behavioral visibility into agent activity.

Compliance Officer

EU AI Act, SOC 2, audit evidence.

- AI inventory and coverage heatmap ready at any time.
- Audit logs and data flow records generated continuously.
- No manual evidence assembly before reviews.

CIO

AI transformation at scale with accountability.

- Cross-account discovery and governance in one platform.
- Closed-loop deployment → enforcement → evidence.
- The posture that lets AI adoption continue.

THE WALLARM ADVANTAGE

Built on the engines that already protect production APIs.

Wallarm's approach to AI governance is grounded in more than a decade of inline API security work. The same detection and enforcement engines that protect production APIs today are now extended to govern the AI workloads those APIs serve.

Wallarm observes and governs AI workloads at the kernel level using eBPF and non-invasive memory analysis. Most AI governance platforms describe what should happen. Wallarm makes it happen, in production, in real time.

TIME TO VALUE

Within minutes

From **helm install** and workload label to active instrumentation.

COVERAGE SCOPE

63 controls

Across nine infrastructure layers and seven security domains.

FOOTPRINT

0 sidecars

Zero code changes. Zero sidecars. Zero instrumentation injected into your workloads.

NEXT STEP

AI is already in production. Governance should be too.

Your AI agents are making decisions, accessing data, and calling external services right now. Wallarm tells you what they're doing, enforces what they shouldn't, and proves it to anyone who asks.

See Wallarm in Action →