LangProtect

AI Everywhere. Security Nowhere.

How Enterprises Can Close the Gap Between AI Adoption and AI Governance

Artificial intelligence has moved faster than any enterprise security strategy anticipated. In two years, AI tools went from controlled pilots to daily infrastructure, used across every function, embedded in products, and connected to internal systems through autonomous agents. The question is no longer whether AI has arrived. It is whether anyone is governing what it does with data.

The evidence suggests most are not. 78% of enterprise leaders call AI governance a top-three priority. Only 31% have a framework in place. In organizations where prompt-level monitoring has been deployed, the findings are consistent, 37% of employee AI prompts contain sensitive data. Most security teams had zero visibility on any of it.

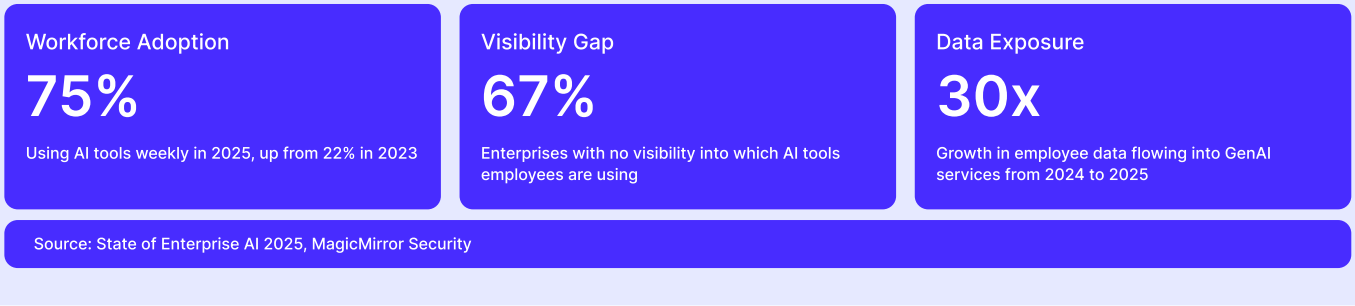
The problem is structural. Enterprise security stacks were built before natural language became a data egress channel. DLP catches files. SIEM logs network events. Neither inspects what an employee types into a prompt window, which is precisely where sensitive data is leaving the organization, at scale, every day.

78% of enterprises say AI governance is a top-three priority.
Only 31% have a framework in place.
67% cannot see which AI tools their employees are using right now

The Gap Is Bigger Than You Think

Your employees did not wait for policy. The day ChatGPT became available, they started using it. The day Copilot launched inside Microsoft 365, it was already open in a hundred browser tabs. AI adoption inside enterprises did not happen through IT procurement. It happened through individual decisions, made quietly, one prompt at a time.

That is not a technology gap, it's a visibility gap. And it is larger than most security leaders realize.



Adoption Without Visibility

Workforce AI adoption jumped from 22% in 2023 to 74% in 2024.

That is not gradual adoption, that is a behavioral shift at organizational scale, happening faster than any governance infrastructure could respond. In a single Guardia deployment, 121 distinct AI applications were auto-discovered across one organization. IT had approved a handful of them. The rest were invisible, each one an ungoverned channel through which employee data was moving every day.

This is what security teams are now calling Shadow AI. It mirrors the early days of cloud adoption, but with higher stakes. When an employee saves a file to a personal Dropbox, the file is the risk. When an employee pastes a patient record into ChatGPT, the conversation, the context, and the inference are all at risk, and none of it triggers a conventional security alert.

What Is Actually Being Sent

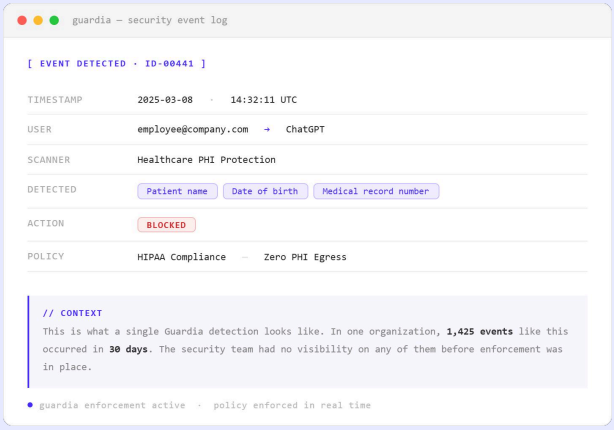
The assumption most enterprises operate under is that employees know better than to include sensitive data in an AI prompt. The data says otherwise.

In one deployment ,we see that:

- 37% of AI prompts contained sensitive data, PII, PHI, source code, credentials, or confidential business content.
- 46% of data-policy violations in early-adopter organizations involved developers pasting proprietary source code directly into AI tools for debugging or code generation.

Finance teams drop customer account data into ChatGPT to draft reports. HR teams paste employee performance reviews into AI tools to write summaries. Not maliciously. Efficiently.

The AI tool receives the data. Process it. Logs it. Potentially uses it. And in most organizations, there is no record the interaction ever happened.



Policy Without Enforcement Is Not Governance

75% of organizations have an AI usage policy on paper. Only 44% have an incident response plan. And almost none have a technical control that enforces the policy at the moment a prompt is submitted.

This is the critical distinction most enterprises are missing. A policy that says "do not share patient data with external AI tools" is not a control. It is a request. Employees working under deadline pressure, trying to move faster, will not pause to cross-reference an acceptable use policy before they type. The control has to be invisible, automatic, and operating at the prompt layer, before the data leaves.

The gap between what organizations believe about their AI security posture and what is actually happening inside their environment is not a small discrepancy. It is the difference between governance and documentation.

WHAT MOST ENTERPRISES ASSUME

- AI usage policy is in place
- Employees follow data handling guidelines
- Security tools would catch a data breach
- Sensitive data stays inside the perimeter

WHAT IS ACTUALLY HAPPENING

- 121 shadow AI apps running undetected
- 37% of prompts contain sensitive data
- DLP, SIEM, and firewalls see none of it
- 162 exposures bypassed warning in 30 days

Three Surfaces. Zero Governance.

Enterprise AI risk does not live in one place. It lives across three distinct surfaces, the employees using AI tools, the applications built on AI models, and the agents operating autonomously on behalf of the business. Most organizations are exposed on all three simultaneously, with different threat profiles on each, and no unified control layer governing any of them.

Surface 01: The Employee Layer

Every employee with a browser has direct, ungoverned access to ChatGPT, Claude, Gemini, DeepSeek, Perplexity, and thousands of other AI applications. They use these tools for real work, drafting documents, debugging code, summarizing reports, and analyzing data. And they include real data in their prompts without considering where that data goes once it leaves the browser.

Shadow AI is rational behavior. Employees adopt the tools that make them productive. The risk is not intent, it is invisibility. When an employee pastes a customer's medical record into ChatGPT to draft a clinical summary, the organization has no record of it, no control over what the AI platform does with that data, and no way to prove to a regulator that it did not happen.

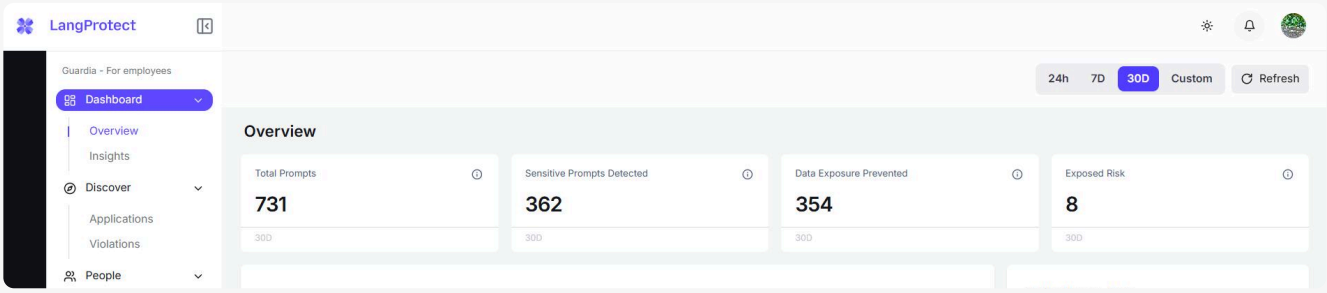
The data from organizations where prompt-level monitoring has been deployed tells a consistent story:

Shadow AI is rational behavior. Employees adopt the tools that make them productive. The risk is not intent, it is invisibility. When an employee pastes a customer's medical record into ChatGPT to draft a clinical summary, the organization has no record of it, no control over what the AI platform does with that data, and no way to prove to a regulator that it did not happen.

The data from organizations where prompt-level monitoring has been deployed tells a consistent story:

- 37% of employee AI prompts contain sensitive data
- 46% of data-policy violations involve developers pasting proprietary source code into AI tools
- 121 distinct AI applications were auto-discovered in a single enterprise deployment — the majority unknown to IT
- ChatGPT accounts for 90% of employee AI usage, followed by Claude at 6%

Each of those applications is an ungoverned data egress channel. Each prompt is a transfer of organizational data to an external system with its own retention policies, training pipelines, and terms of service.



One organization. Thirty days. What Guardia makes visible from day one.

Surface 02: The AI Application Layer

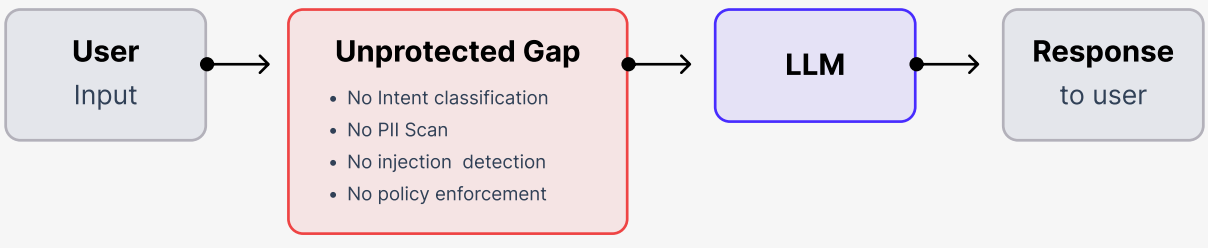
Every AI-powered product an organization ships(customer chatbots, internal copilots, support automation, RAG pipelines), has an unprotected input window between what a user types and what the LLM receives. That window is the primary attack surface for prompt injection, jailbreaking, and data exfiltration through AI applications.

The threats enterprises face at this layer include:

- **Direct prompt injection**: a user crafts an input designed to override the model's system instructions and make it behave outside its defined role
- **Indirect prompt injection**: malicious instructions embedded in documents, emails, or web content that the AI retrieves and processes as trusted input
- **PII leakage in outputs**: the model surfaces sensitive data from its context window or retrieval layer in its response to a user who should not have access to it
- **Jailbreaking**: multi-step manipulation that gradually shifts the model away from its guardrails through seemingly innocuous conversation

72% of organizations have experienced manipulation attempts on their AI applications in the past year.

The majority detected them after the fact, through user complaints, incident reports, or external disclosure, not through runtime monitoring.



Every AI application has this gap. Most organizations have nothing in it.

Surface 03: The AI Agent Layer

AI agents represent the newest and least governed surface. An agent connected via MCP to a company's CRM, code repository, internal knowledge base, or financial system accepts natural language instructions and executes actions autonomously, often across multiple tools in a single workflow, without a human reviewing each step.

The governance challenge at this layer is fundamentally different from the first two surfaces:

- Agents take actions, not just generate responses, a compromised agent can delete records, exfiltrate data, or send communications on behalf of the organization
- Indirect prompt injection through trusted external sources is the primary attack vector, the agent receives a document, email, or web page containing malicious instructions that override its original task
- Tool call chains are opaque, a single agent workflow may invoke five or six tools, with sensitive data flowing through each connection in ways that no existing security tool was designed to trace
- Agent identity is rarely enforced, most organizations cannot attribute agent actions back to a human identity, creating an accountability gap that regulators are beginning to scrutinize

62% of organizations are already experimenting with AI agents.

Almost none have a runtime enforcement layer governing what those agents can access, what data they can share, or whether the instructions they are executing originated from a legitimate source

Why Your Current Security Stack Can't See This

Traditional security tools were not built for AI interactions, not because they are inadequate, but because the threat surface did not exist when they were designed. The result is a precise and dangerous blind spot at the prompt layer.

DLP tools detect sensitive data in files, emails, and structured data transfers. They do not parse natural language. An employee copying a patient record into a Word document triggers DLP. The same employee pasting that record into a ChatGPT prompt does not. SIEM platforms log network events and system activity.

They can confirm an employee connected to chatgpt.com, they cannot tell you what was inside the prompt. Firewalls control traffic destinations, offering only a binary choice: block the AI tool entirely and lose the productivity benefit, or allow it and lose all visibility into what is being shared.

None of these tools were built to inspect a prompt. None of them operate at the browser layer where AI interactions happen. And none of them produce the audit evidence regulators are beginning to require for AI data flows.

Security Tool	What It Secures	Primary Use Case
DLP	**Files, emails, structured transfers**	Natural language prompts
SIEM	**Network events, system logs**	Prompt content, data type
Firewall	**Traffic destinations**	What the employee sent
LangProtect	**Every prompt, every surface, in real time**	Nothing

From Governance Gap to Governance in Action

Governance is not a policy document filed in a SharePoint folder. It is a technical enforcement layer that operates at the moment a prompt is submitted, the moment an API call is made, and the moment an agent takes an action with the audit evidence to prove every decision it made. Four capabilities define what that enforcement layer must do. LangProtect delivers all four, across all three ungoverned surfaces, from a single platform.

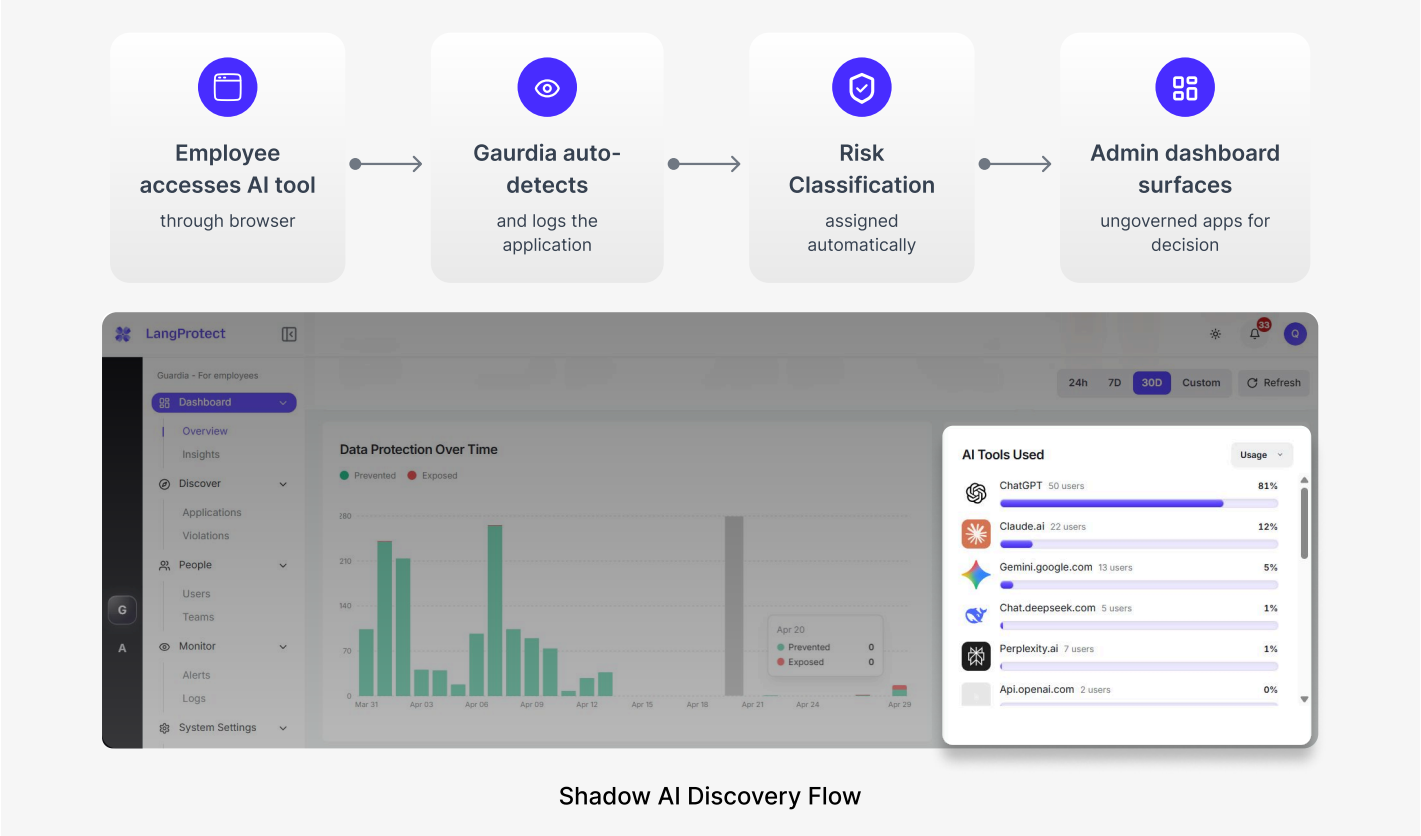
Capability 1: Shadow AI Discovery

Before any organization can govern AI, it needs a complete and accurate inventory of the AI it is actually using not what's approved, but what employees are actively accessing through their browsers every day.

Effective Shadow AI discovery must:

- Automatically detect every AI application accessed across the organization without requiring manual input or employee self-reporting
- Assign a risk classification to each application based on data handling practices, compliance certifications, and geographic data residency
- Surface ungoverned applications to administrators in real time with the context needed to make an approve, monitor, or block decision
- Track usage patterns over time, which tools are growing, which teams are using them, and which are accessing high-risk AI applications repeatedly

LangProtect Guardia executes this automatically from the first day of deployment. In one enterprise deployment, one hundred and twenty-one distinct AI applications were discovered in thirty days, the majority of which IT had never approved or catalogued. Each one was an unmonitored data egress channel operating inside the organization's perimeter.



Shadow AI Discovery Flow

Capability 2: Real-Time Data Protection at the Prompt Layer

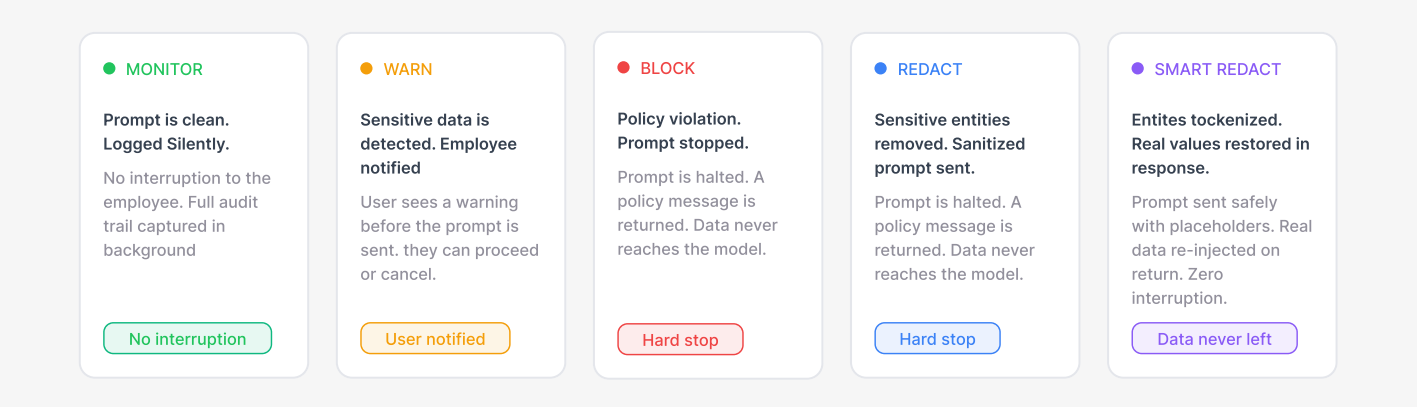
Every prompt submitted to any AI tool must be scanned before it is sent, not logged after the fact, not reviewed in a weekly report. Scanned in real time, at the moment of submission, before the data reaches the model.

Guardia's scanner engine runs thirty-two built-in scanners in parallel across every prompt, covering:

- PII, PHI, and PCI detection across eighteen countries and all eighteen HIPAA Safe Harbor identifiers
- Source code and credentials protection for developer workflows, where 46% of enterprise data violations originate
- Prompt injection and jailbreak detection to prevent policy circumvention
- Intent classification using a self-hosted LLM - employee prompts are never sent to an external AI provider for scanning
- Custom scanner creation for organization-specific data types, terminology, and compliance requirements

Enforcement operates through five outcomes, Monitor, Warn, Block, Redact, and Smart Redact. Smart Redact is the most powerful: sensitive entities are tokenized before the prompt leaves the browser, the sanitized prompt is sent to the AI model, and the real values are restored in the response before the employee sees it. The employee receives a complete, contextually accurate AI response. The data never left the organization.

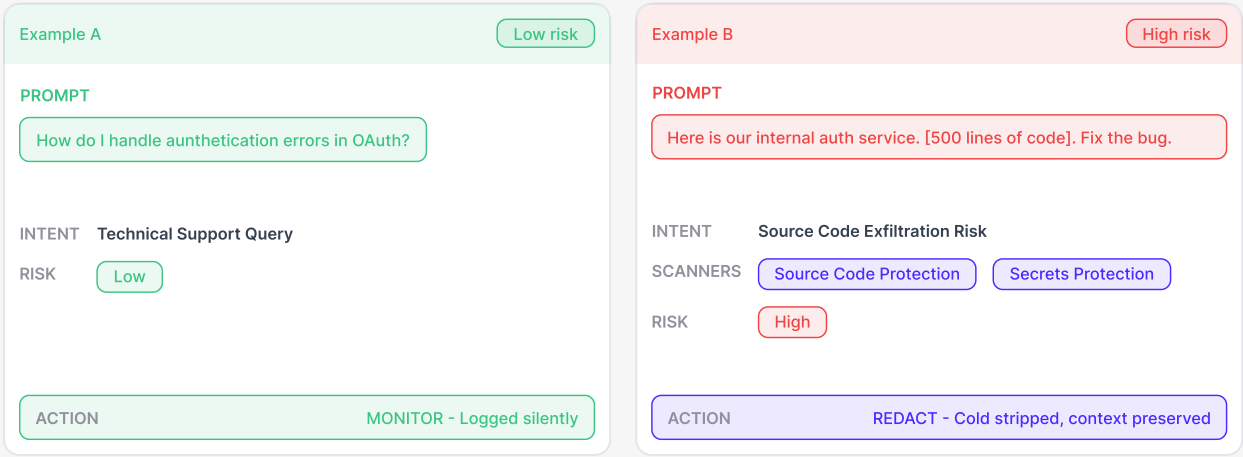
For AI applications, Armor applies the same enforcement model at the API layer, scanning every user input and every model response in under fifty milliseconds, with no measurable impact on user experience.



Capability 3: Intent Classification Beyond Keywords

Pattern matching fails against natural language because the same words carry entirely different risks depending on context. A developer asking how to handle an authentication error and a developer pasting five hundred lines of proprietary source code into the same AI tool look identical to a keyword filter, and completely different to an intent engine.

Guardia's intent classification operates at the semantic level, understanding what an employee is trying to accomplish rather than what words they used to accomplish it. This allows the enforcement layer to detect adversarial prompts specifically designed to avoid triggering keyword-based controls, jailbreak attempts, instruction overrides, and role-play manipulations, as well as high-risk workflows that use entirely legitimate language but involve sensitive data in the context of the request.



Both prompts contain the word "code." Only intent classification tells them apart.

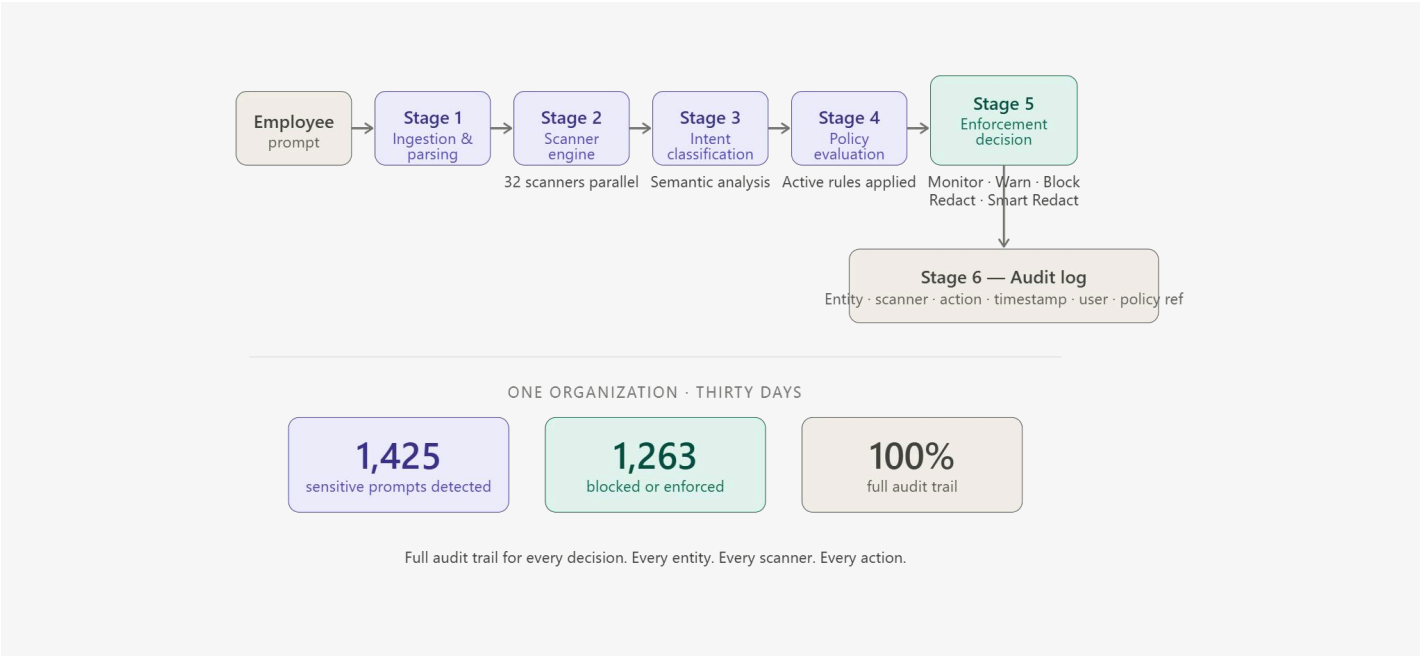
Capability 4: Policy Enforcement With Audit Evidence

Enforcement without evidence is insufficient for any regulated enterprise. When a HIPAA auditor, a GDPR data protection authority, or an internal compliance team asks whether PHI reached an external AI model, the answer must be a documented, timestamped, exportable record.

Every LangProtect enforcement decision produces a log entry containing:

- The specific entity types detected and the scanner that identified them
- The enforcement action taken and the policy rule that triggered it
- The user identity, the AI tool accessed, and the timestamp of the interaction
- A unique audit ID for cross-referencing with SIEM, HRMS, or compliance management systems

For organizations operating under HIPAA, GDPR, SOC 2, or India's DPDP Act, this audit trail is the difference between a governance program that satisfies a regulator and one that only satisfies the internal team that built it.



The Window to Act Is Now

The governance gap documented in this whitepaper is not a projection. It is a present reality, measurable today in every organization where employee AI usage is monitored at the prompt layer.37% of prompts contain sensitive data. One hundred and twenty-one shadow AI applications running without IT knowledge. One hundred and sixty-two data exposures that bypassed warning in a single thirty-day period. These are not worst-case scenarios, they are baseline findings from organizations that chose to look.

The enterprises that will scale AI with confidence are not the ones that move slowest. They are the ones that build governance infrastructure in parallel with adoption, so that every new AI tool, every new use case, and every new agent deployment lands inside a controlled, auditable, enforcement-ready environment rather than expanding an already ungoverned surface.

Three things determine whether an organization closes this gap or continues to widen it. First, visibility, knowing what AI tools are in use, which teams are using them, and what data is moving through them. Second, enforcement, a technical control operating at the prompt layer that acts before data leaves, not after. Third, evidence, an audit trail that satisfies regulators, supports compliance reviews, and gives security leadership a defensible answer when the question comes.

All three are available today. The cost of not having them is not theoretical. It is measured in undetected PHI exposures, unreported source code leakage, and compliance findings that could have been prevented by a browser extension deployed in an afternoon.

Get to know How LangProtect Detect a Threat

See LangProtect intercept sensitive data in a real prompt, in real time, across your AI tools, in your environment.

BOOK A DEMO >

ABOUT LANGPROTECT

LangProtect is an enterprise AI security platform built to secure the full spectrum of organizational AI activity across employees, applications, and autonomous agents.

Guardia	Armor	Vector
Latency Employee AI interactions	Latency AI applications and LLM APIs	Latency AI agents and MCP connections
Primary Use Case Shadow AI discovery, data leakage prevention, compliance audit trail	Primary Use Case Prompt injection defense, PII redaction, runtime policy enforcement	Primary Use Case Agentic workflow governance, unauthorized tool call prevention

LangProtect deploys at the browser and API layer, no network reconfiguration, no endpoint agent complexity, no disruption to existing workflows. Detection runs in under 50ms. Audit logs are available from the moment of deployment.