Forcepoint

Datasheet

# Forcepoint AI Security

### What It Governs:

- Data Protection Across AI Interactions
- Shadow AI Discovery and Control
- Data Access Controls for Sanctioned AI
- Custom AI Agent Monitoring and Governance
- AI Agent Gateway: Runtime Control and Enforcement

### Key Benefits:

- **AI Governance Dashboard:** Single view across every sanctioned AI platform, shadow AI tool and agents.
- **Zero Reclassification:** Existing DLP policies extend to every AI channel on day one. No policy rebuilds, no new classifiers.
- **Complete Coverage:** One platform governs employee prompts, shadow AI tools, custom agents and cloud-resident agents.
- **Built-In Compliance:** Every interaction and agent action is logged with full attribution for security, compliance, audit and forensic investigation.
- **Immediate Time to Value:** Forcepoint Adaptive Risk Intelligence Assistant (ARIA) guides administrators from first login to a governed environment.

## Overview

Most enterprise security programs were built around a straightforward assumption: sensitive data moves through channels the security team can monitor. That assumption does not hold in today's AI environment. When an employee pastes a customer record into ChatGPT or asks Copilot to summarize files in SharePoint, none of that activity passes through email filters, web proxies or endpoint controls. Shadow AI tools compound the effect further: employees are using AI applications the security team has never reviewed, and every unsanctioned tool is a data exposure with no visibility and no audit trail.

AI agents add to this risk environment. A custom agent can chain dozens of tool calls, query sensitive data stores and write to external systems across a single session, entirely outside the visibility of every security tool currently in place. These are not theoretical risks. They are how AI is being used in production environments today.

Forcepoint AI Security closes that gap. It governs every AI interaction across the enterprise, from the prompt an employee types to the tool call an autonomous agent makes at machine speed, with the same enforcement controls already protecting every other data channel: SaaS, web, email and endpoints. Unified policies enforced from a data-centric platform that classifies sensitive data, governs every interaction and stops every risk automatically.

## AI Detection and Response

API-based monitoring and response across sanctioned AI platforms and custom agents

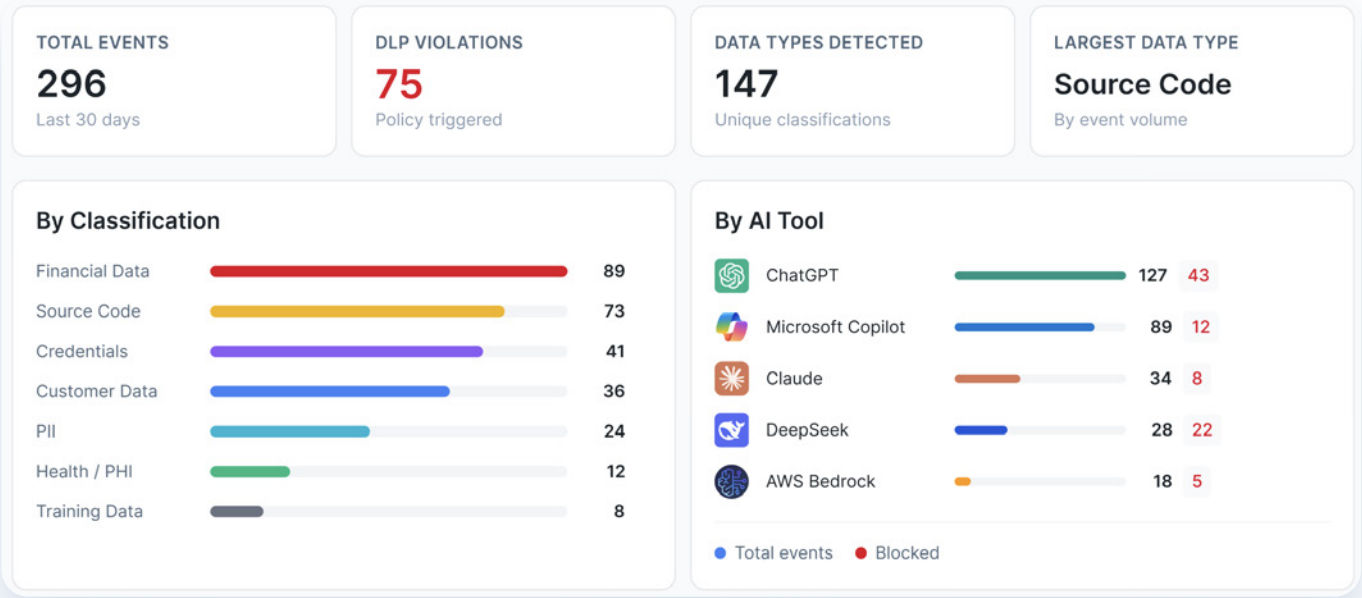
### Data Protection Across AI Interactions

### Do you know what your employees are putting into ChatGPT?

Every AI prompt is a data transfer. Employees routinely paste customer records, financial data and source code into AI tools, moving sensitive information outside enterprise control with no interception and no record. Existing DLP controls were not designed to inspect AI prompt and response payloads.

Forcepoint AI Security brings AI interactions inside the same policy framework governing every other data channel. Existing classification policies apply to prompts, responses and file uploads across ChatGPT Enterprise, Microsoft 365 Copilot, Claude for Enterprise and AWS Bedrock. Policy violations trigger a remediation action: audit, quarantine, copy or delete.

- → **Zero reclassification:** For organizations already running Forcepoint DLP, existing policies extend to AI channels on day one. No new taxonomy required.
- → **Prompt and response inspection:** Detection covers typed input, pasted content, structured uploads and AI-generated responses, not just file attachments.
- → **Retroactive discovery:** Historical backfill surfaces sensitive data events that were invisible before the product was deployed.
- → **No workflow disruption:** API-based scanning requires no changes to how employees work and adds no overhead to the sanctioned tools they use.



## Data Access Controls for Sanctioned AI

### "Copilot is reading your SharePoint files right now. Do you know what it can access?"

Microsoft Copilot has direct access to every file in OneDrive and SharePoint that an employee can reach. Enterprise permissions were designed for humans, not AI assistants that can scan and summarize thousands of files in seconds. Overshared or misconfigured files instantly surface HR records, financial data, board materials and intellectual property to users who were never meant to see them.

Forcepoint AI Security uses existing classification intelligence to understand what sanctioned AI assistants can access before any prompt is typed. MIP tagging prevents Copilot from summarizing or surfacing confidential files. API-based remediation revokes access to overshared files before they can be ingested. The same connector framework extends full session audit to ChatGPT Enterprise and Claude for Enterprise.

- **Pre-access classification:** Sensitive files are identified before any AI assistant interaction begins. Controls are applied upstream of retrieval, not after exposure.
- **MIP tagging for Copilot:** Confidential files in OneDrive and SharePoint are tagged to prevent Copilot from accessing, summarizing, or surfacing them in prompt responses.
- **Automatic remediation of oversharing:** API-based remediation revokes access to overshared content without requiring manual permission review.
- **Full prompt and response audit:** Every session on a connected AI platform is logged and attributed, covering what was accessed, what was asked and what the platform returned.

## Custom AI Agent Monitoring and Governance

### "Can you show me every custom AI agent running in your environment right now?"

Enterprises are deploying custom AI agents on AWS Bedrock and Copilot Studio with access to internal data sources. Most organizations have no inventory of what agents are running, what data they can reach or whether any are operating outside intended boundaries. When something goes wrong, there is no audit trail and no attribution.

Forcepoint AI Security ingests and parses invocation logs from AWS Bedrock via CloudWatch and S3, reconstructing session traces and attributing activity to identities via IAM, Okta or Entra ID. Every agent is scored for risk based on data sensitivity. Security teams can correlate agent actions against DLP events and act before an incident occurs.

- → **Full attribution:** Every agent action is tied to both the agent identity and the human who triggered it. That linkage is what makes AI incident investigation possible.
- → **Early warning on agent sprawl:** Automated alerts fire when new agents are published or when existing agents add new external connections.
- → **Actionable governance controls:** Pause a run, require human approval, revoke access or quarantine agent-written files.
- → **Compliance-ready audit logs:** Agent activity records support EU AI Act Article 12, GDPR Article 30, DORA, NIS2 and SEC AI disclosure requirements.

| Agent | Owner / Team | Profile | Scope | Token | Last Active |
|---|---|---|---|---|---|
| **Sale-Copilot-Prod**<br>cl_748651dfbd...sdgd | Mary Rodriguez<br>Revenue Ops | Read-Only | 6 tools | • Active | 18-May-26<br>11:11:11 |
| **Market-Analyzer-Prod**<br>cl_853462dfbf...ghij | John Smith<br>Marketing Lead | Editable | 4 tools | • Expiring | 20-May-26<br>12:15:30 |
| **Customer-Insights-Prod**<br>cl_948273dfbe...ijkl | Emma Johnson<br>Data Analyst | Read-Only | 5 tools | • Active | 22-May-26<br>14:30:45 |
| **SalesTracker-Prod**<br>cl_748652dcba...mnop | Liam Brown<br>Sales Manager | Editable | 3 tools | • Inactive | 23-May-26<br>09:00:00 |
| **Product-Feedback-Prod**<br>cl_652348dfgh...qrst | Olivia Taylor<br>UX Designer | Editable | 2 tools | • Active | 24-May-26<br>10:05:12 |
| **Competitor-Tracker-Prod**<br>cl_763291dfgh...uvwx | Noah Wilson<br>Research Analyst | Read-Only | 6 tools | • Active | 25-May-26<br>13:45:20 |
| **Campaign-Optimizer-Prod**<br>cl_253471dfgh...yzab | Ava Lee<br>Marketing Specialist | Editable | 5 tools | • Inactive | 26-May-26<br>15:30:00 |
| **Lead-Management-Prod**<br>cl_847162dfgh...cdef | Sophia White<br>Lead Strategist | Read-Only | 7 tools | • Active | 27-May-26<br>11:20:33 |

## Shadow AI Discovery and Control

Discovery, risk scoring and access control for every unsanctioned AI tool in the enterprise

### "How many AI tools are your employees using that you did not approve?"

Shadow AI tools proliferate at the speed of a browser extension. Employees adopt personal AI accounts, writing assistants, coding tools and document processors without going through security review, and organizations now manage an average of 37 AI agents while many security teams lack visibility into what those agents access. Every unsanctioned tool is a potential data exposure existing controls cannot see.

Forcepoint AI Security gives security teams a continuously updated, risk-scored inventory of every unsanctioned AI application in use. Security teams see full activity detail per user and per tool: visits, prompts, responses, file uploads and file downloads. When employees use personal rather than sanctioned corporate accounts, AI Security detects the distinction and applies policy accordingly.

- → **Risk-scored AI inventory:** Every unsanctioned AI application is categorized and scored, grounding access decisions in actual risk.
- → **Activity drill-down:** Full visibility into what employees are doing across every shadow AI tool, including prompt activity and file transfers.
- → **Corporate versus personal account detection:** Identify whether employees are using AI tools through sanctioned corporate accounts or personal accounts, and enforce policy accordingly.
- → **Access control per tool:** Allow, block or restrict uploads and downloads per AI application, scoped by user or department.

**Unsanctioned AI Apps**

| App Name | Category | Risk | Users | Trend |
|---|---|---|---|---|
| **Google Gemini** gemini.google.com | Generative AI - Conversation | Low | 89 | |
| **Perplexity AI** perplexity.ai | Generative AI - Conversation | Low | 87 | |
| **Midjourney** midjourney.com | Generative AI - Multimodal | Medium | 43 | |
| **DeepSeek** deepseek.com | Generative AI - Conversation | High | 31 | |
| **Hugging Face** huggingface.co | Other - AI App | Medium | 29 | |
| **Poe** poe.com | Generative AI - Conversation | Low | 22 | |
| **Jasper AI** jasper.ai | Generative AI - Text and Code | Low | 18 | |
| **Cursor** cursor.so | Generative AI - Text and Code | High | 15 | |

## AI Agent Gateway

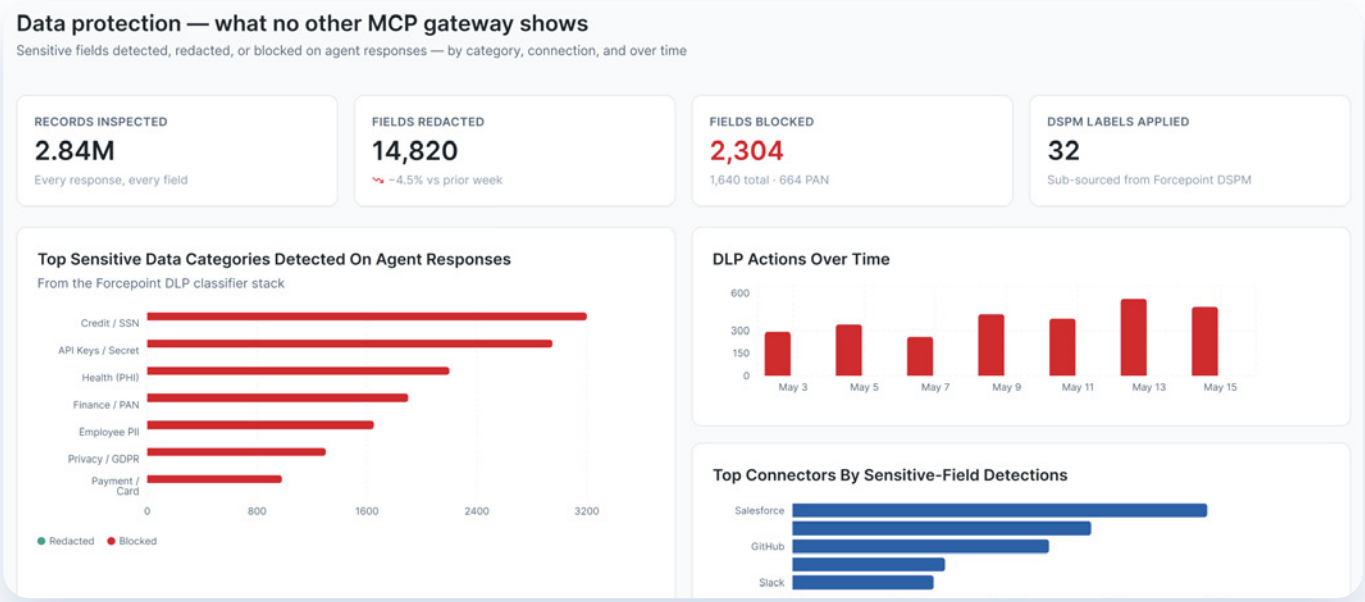
MCP-native inline enforcement between cloud-resident AI agents and the SaaS applications they call

### "Can you stop an AI agent from accessing data it should not touch?"

Cloud-resident agents connect directly to SaaS applications, such as Microsoft 365, Salesforce and Jira, operating with the full permissions of the service account they were handed at deployment. There is no mechanism to limit what data an agent reads, block fields it should never see or halt an operation mid-execution. A compromised agent is effectively a compromised credential with unrestricted application access and no accountability.

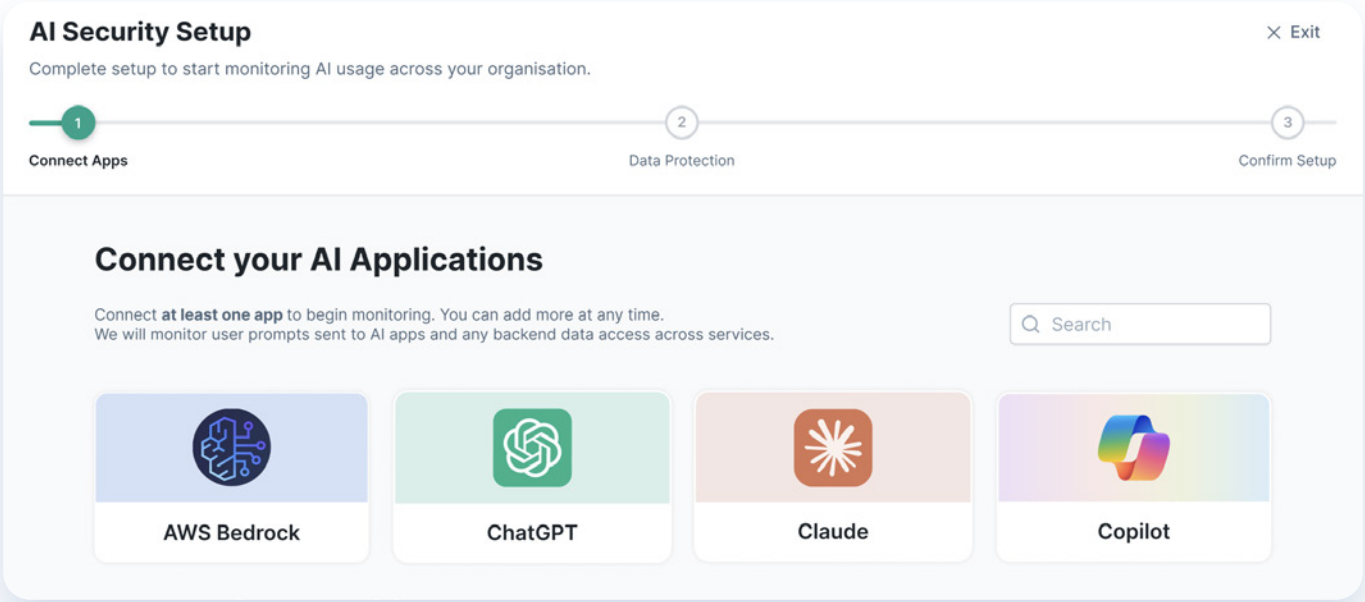
Forcepoint AI Agent Gateway is a purpose-built MCP gateway sitting as the mandatory intermediary between cloud-resident AI agents and the SaaS applications they call. No agent ever holds credentials to any business application directly. Each agent operates within a defined scope of approved tools, issued short-lived tokens subject to instant revocation. Every response is inspected at field level before the agent can read it: PII, source code, credentials and secrets are blocked or redacted before transmission. Write and delete operations can be held for human approval via Teams, Slack or email. Every transaction is immutably logged.

- → **Agents never hold direct credentials:** The Gateway brokers all application credentials. A compromised agent cannot authenticate to any connected application directly.
- → **Least privilege enforced by design:** Each agent is registered with a defined scope of approved tools and issued short-lived tokens. Agents cannot reach data or operations outside their approved boundary.
- → **Field-level data protection:** PII, source code, credentials and secrets are blocked or redacted before the agent can read, aggregate or transmit them.
- → **Human-in-the-loop controls:** Write and delete operations can be held for human approval. The approver sees agent name, user identity, tool, parameters and data classification before deciding.
- → **Anomalous behavior detection:** Unusual record-volume queries are detected and stopped at the execution layer before bulk data extraction can complete.
- → **Immutable audit trail:** Every transaction generates a complete record: agent identity, user identity, tool called, data classification, policy decision, fields blocked or redacted, record count and timestamp.



## Fast Time to Value

Forcepoint AI Security is designed to deliver value from day one. ARIA, Forcepoint's AI-powered onboarding assistant, launches automatically on first login and guides the administrator through connector setup and policy configuration, recommending a starting DLP policy set based on the organization's region, industry and data sensitivity profile. No professional services required. Historical interaction data is backfilled automatically on first connection, so the AI Governance Dashboard has real content from the moment setup is complete.

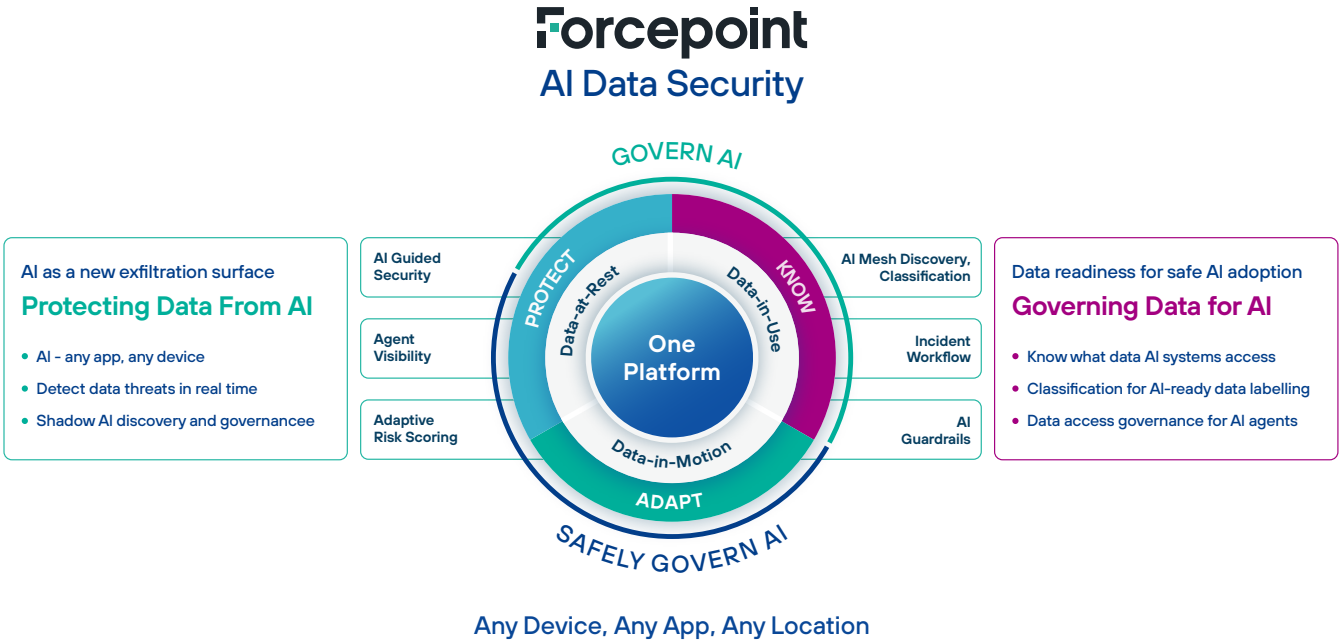


## Key Capabilities

| FEATURE | BENEFIT |
|---|---|
| **API-based connectors** | Deep visibility and control over ChatGPT Enterprise, Microsoft 365 Copilot, Claude for Enterprise and AWS Bedrock. |
| **DLP policy inheritance** | Existing Forcepoint DLP classifications extend automatically to AI prompts, responses and file uploads without requiring a new taxonomy. |
| **Prompt and response logging** | Complete text of every prompt and AI response captured, indexed and searchable by user, date range, classification label and AI tool. |
| **File operation visibility** | Every file uploaded to a sanctioned AI tool is detected, classified against the organization's DLP taxonomy and logged with full audit context. |
| **AI Governance Dashboard** | Single real-time view across all sanctioned AI platforms, shadow AI tools and agent activity. KPI tiles, trend data, risk-ranked findings and guided next actions in one operational view. |
| **Agent inventory and governance** | Live inventory of AI agents across Copilot Studio, ChatGPT Enterprise and AWS Bedrock, with agent identity, authentication mode, connected data sources and risk score surfaced per agent. |
| **Shadow AI discovery and control** | Continuously updated inventory of unsanctioned AI tools with risk scores, full activity detail by user and tool, corporate versus personal account detection and access control policies. |
| **Data access guardrails** | Classification-driven identification of overshared or mislabeled files before AI assistants can access them. MIP tagging and API-based remediation available for Microsoft 365, SharePoint, OneDrive and Microsoft Copilot. |
| **Remediation actions** | Allow, allow and monitor, restrict action, delete conversation history, revoke file access, quarantine agent-written files, pause agent run or require human approval. Available actions vary by interaction type and enforcement surface. |
| **ARIA-guided setup** | AI-powered onboarding assistant guides administrators through connector setup and policy configuration, recommending a starting policy set based on region, industry and data sensitivity. No professional services required. |
| **AI Agent Gateway** | Mandatory intermediary between cloud-resident AI agents and connected SaaS business applications. Agents never hold direct application credentials. Field-level data protection, configurable human-in-the-loop approvals and immutable audit logging on every agent-to-application transaction. |
| **Compliance audit export** | Exportable audit log covering all AI interactions, agent actions and policy events. Supports EU AI Act, NIST AI RMF, GDPR Article 30, DORA, NIS2 and SEC AI disclosure requirements. Exportable in CSV, JSON and PDF. |

Forcepoint AI Security is built on the Forcepoint AI Data Security platform, which classifies sensitive data before AI touches it and enforces policy at every interaction across the enterprise. One policy framework governs email, web, endpoint, cloud applications and AI from a single control plane.

The platform addresses two sides of the same problem: protecting enterprise data from AI tools and governing data so AI can operate safely. Discovery, enforcement and risk insights run on the same engine, giving security teams a single operational view across every surface AI touches.






Forcepoint.com/contact