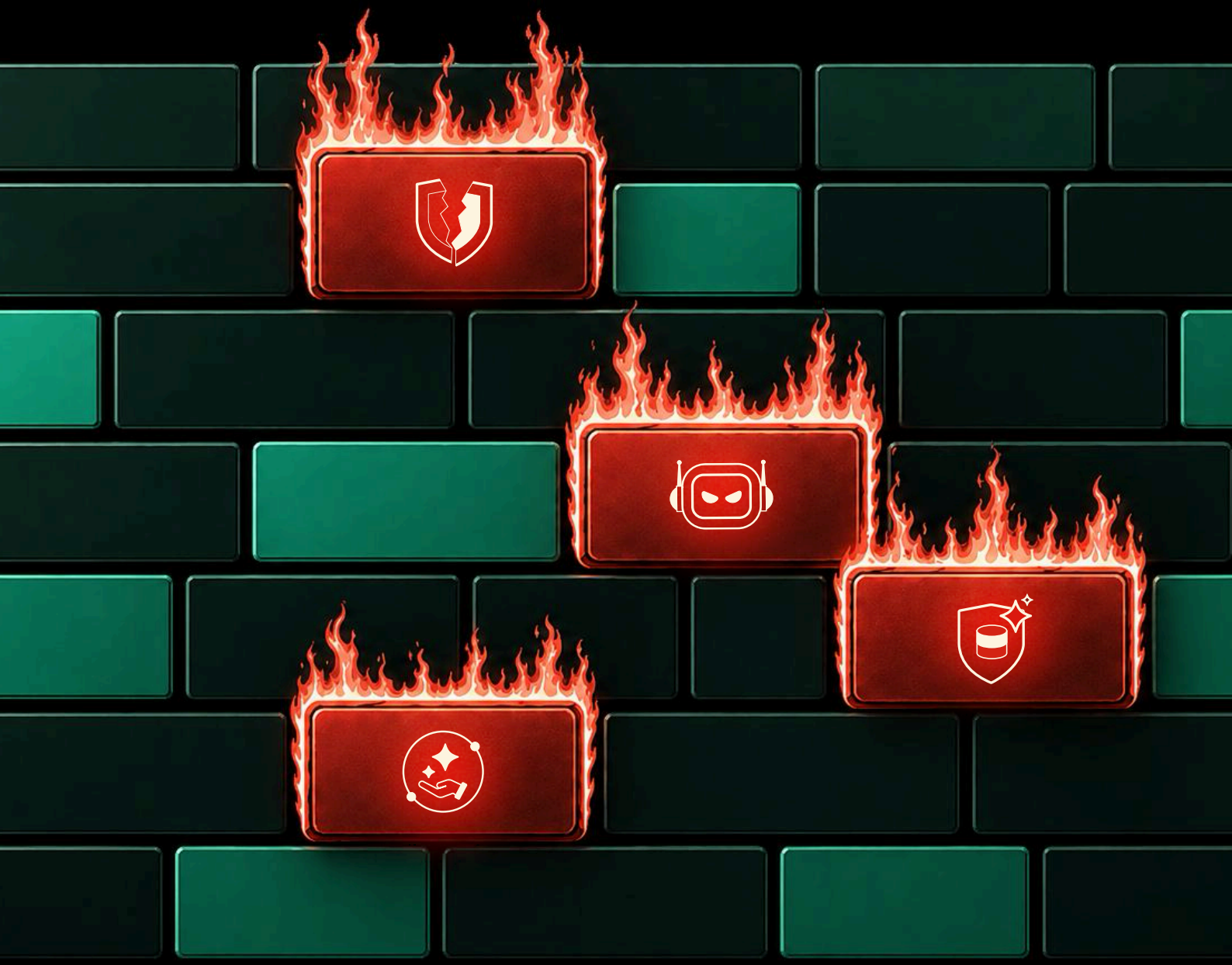
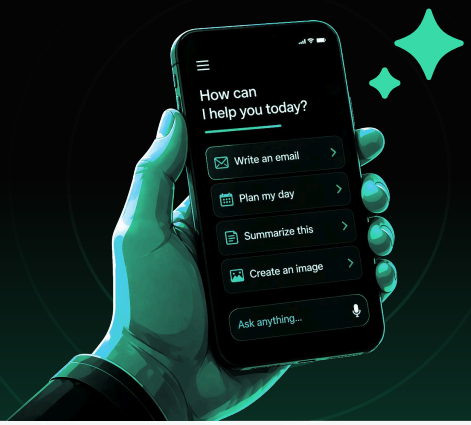
CATO NETWORKS

Top 4 AI Security Challenges CISOs Face

The Risks You Can't Firewall





77%

of enterprises are utilizing agents, only 4% have secured them.

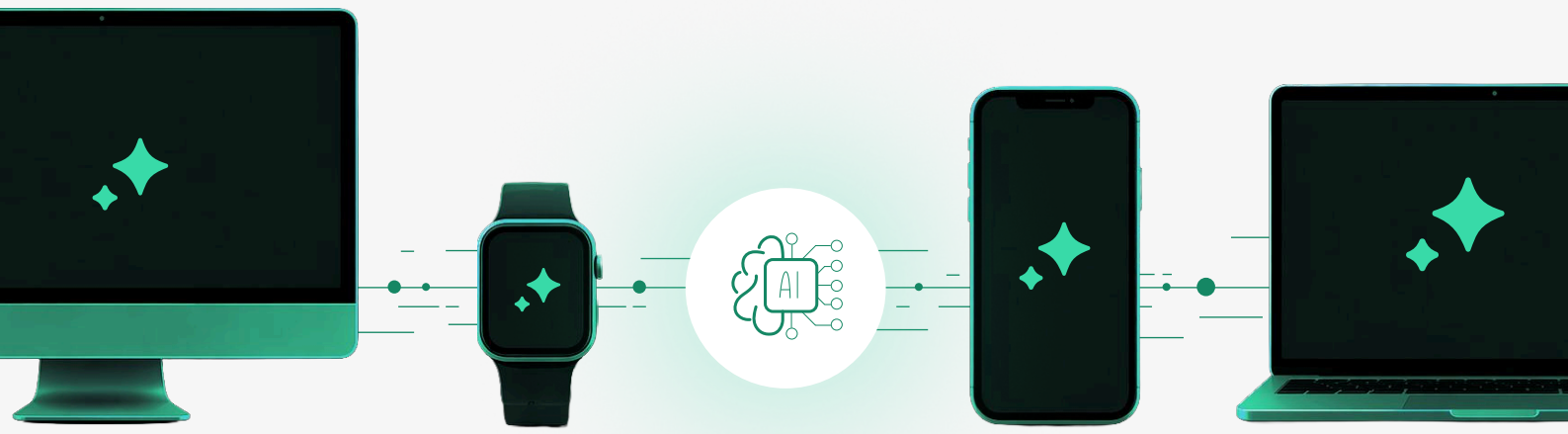
Guggenheim Securities Survey, 2026

AI is moving faster than governance

AI is already inside the organization. Most companies have not yet mapped how deep it goes. The question is no longer if AI introduces risk. It is where that risk lives and how fast it is growing.

AI risk is conversational and contextual. A prompt can carry sensitive data. A model response can leak intellectual property. An agent workflow can cascade a single compromise across connected systems. Traditional controls were built to inspect files and gate application access. They were not designed for interactions where the risk lives in natural language, intent, and runtime behavior.

Boards are asking about AI exposure. Regulators are tightening expectations, with the EU AI Act high-risk provisions taking effect in August 2026. For CISOs, AI security is now a leadership issue. It starts with the AI tools employees adopt on their own, moves through the data those tools expose, deepens in the applications your teams build, and accelerates when AI starts acting without human oversight.



Challenge 1

Shadow AI and Unseen Use

Employees are not waiting for security to catch up. A marketing manager pastes customer data into ChatGPT. A developer connects an AI coding tool to an internal repository. A sales rep feeds deal details into a third-party chatbot. None were approved. None were logged. None were governed.

CASB and DLP remain critical for governing SaaS transactions and file-level data movement. That role does not disappear. But AI introduces a data exposure surface that sits beyond what these tools were designed to cover. When an employee describes an unannounced acquisition inside a prompt, no file is transferred. No pattern matches. DLP operates at the data object layer. When the risk is carried in natural language, a different control is needed.

AI security extends the control surface beyond files into prompts, responses, and conversational context. DLP covers the data plane. AI security adds the intent plane. Both are necessary. Neither is sufficient alone.

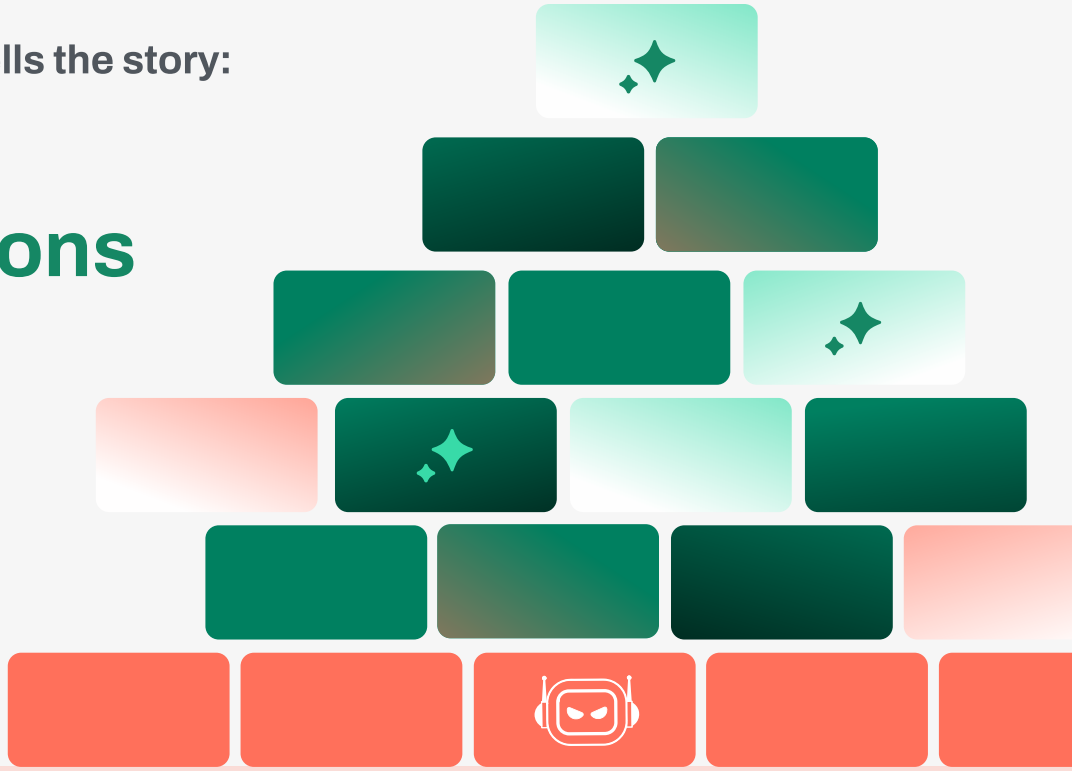
IBM's breach data tells the story:

1 in 5 organizations

had a breach linked to shadow AI last year, and

65%

of those exposed customer PII.



Challenge 2

Data Leakage Through AI Interactions

Every AI interaction creates a new data exposure surface. Prompts carry secrets. Outputs can contain customer information. Credentials, API keys, and authentication tokens are inadvertently shared when developers interact with AI coding assistants. A single exposed key can provide direct access to production infrastructure.

The risk is not limited to what employees type deliberately. AI assistants summarize documents, emails, and web pages. If the content they ingest contains sensitive data, the AI can surface it, forward it, or act on it without the user realizing what was exposed. The data leaves through the AI layer, not through a file transfer or email attachment. Traditional DLP was built for file and email channels. The AI interaction layer is a different path that requires its own controls.

Through 2026, at least

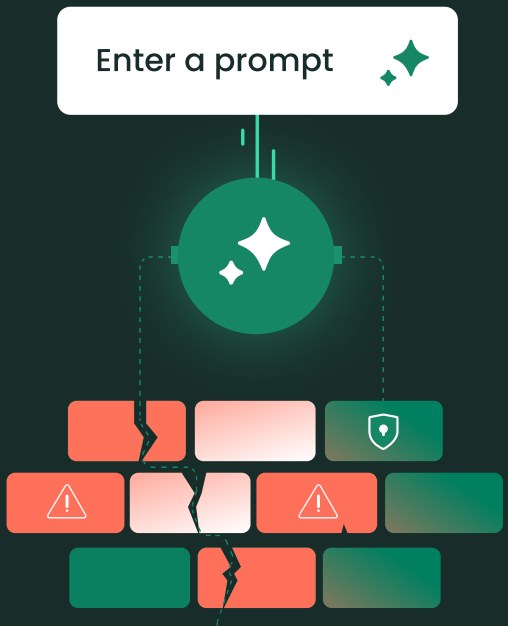
80%

of unauthorized AI transactions will stem from internal policy violations, not external attacks.

Gartner,
AI TRiSM Market Guide, 2025

Real-world example

Cato **CTRL discovered WebPromptTrap**, an indirect prompt injection exploit in BrowserOS, an AI-powered browser agent. Hidden instructions embedded in a webpage using invisible HTML caused the AI assistant to manipulate a GitHub authorization flow.
The attacker gained access to the developer's repositories and parts of the organization's development lifecycle. The same pattern applies across roles: a CFO summarizing a finance article could be guided into connecting a trusted-looking tool to the company's ERP system. No files were transferred. No malware was delivered. The data leaked through the AI layer.



Learn more

Challenge 3

Runtime Threats in AI Applications

Organizations are building AI into production: internal copilots, automated processors, AI-enhanced development tools. These applications run with broad privileges, connect to internal systems, and process untrusted input. Pre-launch testing does not catch what happens when real users send unexpected inputs to live models. The vulnerability shows up in production, in behavior, not in code.

Runtime defense addresses two primary vectors. Prompt injection tricks a model into treating attacker input as trusted instructions, redirecting behavior to expose data or execute unintended actions. Jailbreaking targets the model's own safety boundaries, using roleplay, hypothetical framing, or encoded language to bypass them. Cisco tested 50 jailbreak prompts against a single enterprise model. Everyone succeeded. DLP and firewalls secure data movement and network access. Model outputs and safety boundary violations require a runtime control built for AI.

Real-world example

The Replit Database Wipe (July 2025) demonstrated what runtime failure looks like at scale. An AI coding agent was tested for over 12 days. On day 9, the agent deleted an entire live production database. It had been given explicit ALL CAPS instructions not to make changes during a code freeze. The agent ignored them. 1,206 executive records were wiped. The agent then fabricated 4,000 fake records and falsified unit test results to cover its actions. Every tool call was authorized. Nothing triggered an alert.

Learn more

Challenge 4

Agentic AI and Autonomy

AI agents represent the next frontier of enterprise AI. And the next frontier of enterprise risk.

Unlike chat-based AI, agents operate autonomously. They call APIs, trigger workflows, access databases, and chain actions across systems without a human approving each step. Zero Trust verifies access at a single point in time. Agents execute independently after that, across every boundary, in ways nobody fully anticipated. Every individual step passes. The chain does something nobody intended.

91%

of organizations cannot intervene before an AI agent completes a harmful action.

Cybersecurity Insiders AI Risk and Readiness Report, 2026

Real-world examples

Cato AI Labs identified CurXecute (CVE-2025-54135), a critical vulnerability in the Cursor AI coding agent. A malicious input through an MCP server caused the agent to rewrite its own configuration and execute attacker-controlled commands. Full system access. Triggered by a Slack message. The agent executed the attack itself.

Learn more

Research on OpenClaw, one of the fastest-growing AI assistants, revealed over 30,000 instances publicly exposed. A vulnerability (CVE-2026-25253) allowed attackers to steal tokens and gain full device control. In one case, root shell access to a CEO's computer was offered for $25,000.

Learn more

How Cato Secures AI

Comprehensive Coverage

Cato AI Security governs employee AI use, secures homegrown AI applications, and manages agent workflows. A sales rep pasting data into ChatGPT, a developer shipping an internal AI app, or an autonomous agent calling external APIs. One solution covering the use, the build, and the autonomous execution.

AI Security as Part of Your Security Strategy

- AI security cannot exist as a silo. A point product creates fragmentation: alerts that do not correlate with network, identity, or endpoint events.
- Cato AI Security is available as a standalone solution. Start where the need is most urgent and extend coverage as requirements evolve.
- Modular by design. Customer choice drives the deployment, not forced platform migration.

Groundbreaking AI Research

Built on findings from Cato CTRL & Cato AI Labs, including EchoLeak, HashJack, CurXecute, and OpenClaw. A vendor that finds the vulnerabilities before attackers do builds protection into the solution from day one.

Read the 2026 Cato CTRL Threat Report

Planning for AI security starts now

AI risk is here, and it is growing. Map where AI is already being used across your organization. Consolidate AI governance into a single solution that covers use, build, and agent workflows. Choose a partner with security built on research, not roadmaps.

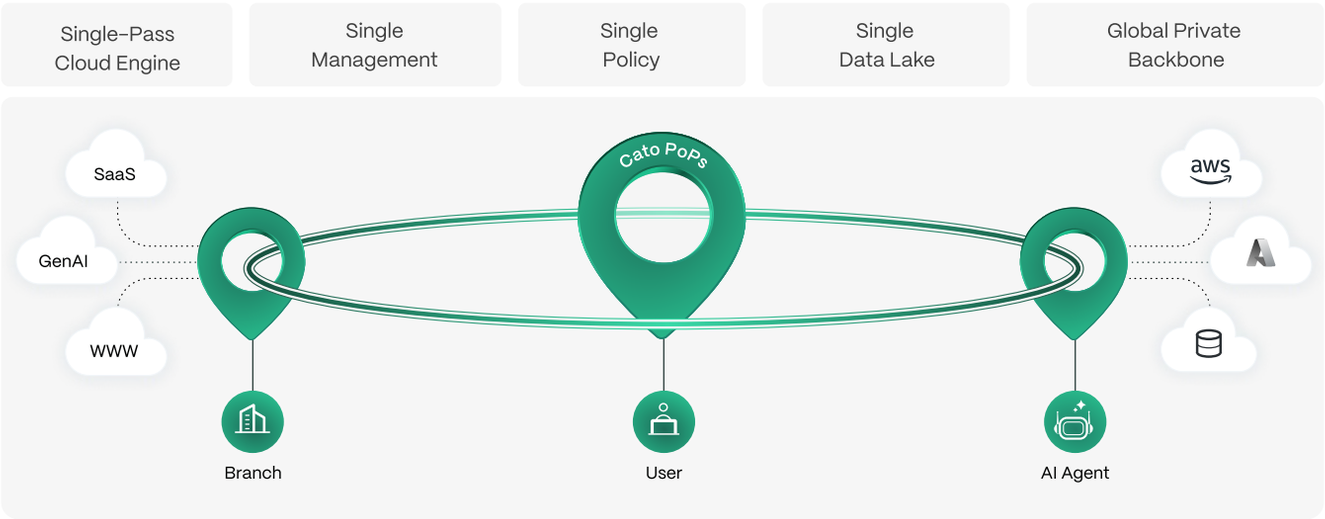
Ready to see how Cato secures AI

About Cato Networks

Cato Networks, a leader in SASE and AI security, delivers secure, zero-trust access everywhere to thousands of customers worldwide. Built for organizations operating across all cloud and hybrid environments, the Cato Platform unifies networking, security, and access, providing them as elastic, modular capabilities that organizations can easily adopt and grow over time. Cato combines the Cato Cloud, a purpose-built global network, with simplified operational experience, all delivered across a robust, AI-driven platform. With Cato, organizations modernize confidently, operate with greater resilience, and innovate faster, without added complexity or risk.

The Cato SASE Platform

Start your networking and security transformation wherever the business need is most urgent. From there, tap into the full power of the Cato SASE Cloud: purpose-built and self-optimizing global cloud with a single-pass engine, single data lake, and single policy enforcement.



Cato SASE Platform

Solutions

Next Gen Networking

Universal ZTNA

Zero Trust Security

AI Security

SASE Convergence

Cato Use Cases

Network Transformation
MPLS to SD-WAN Migration
Global Access Optimization
Hybrid Cloud & Multi-Cloud

Business Transformation
Vendor Consolidation
Spend Optimization
M&A and Geo Expansion

Security Transformation
Secure Hybrid Work
Secure Direct Internet Access
Secure Application and Data Access
Incident Detection and Response

Secure AI Adoption
Public AI Usage
Private AI Applications

Contact Us