eBook

MORPHISEC

The Ultimate Guide to AI Usage Control

Governing shadow AI and autonomous agents on the endpoint

A practical guide for security leaders securing the autonomous enterprise.

What's Inside This Guide

AI is now on every endpoint – in coding assistants, desktop apps, local LLMs, browser extensions, and autonomous agents. Most of it is invisible to the tools you already own, and none of it is governed by firewalls, identity, or EDR. This guide shows you how to take back control: to discover the AI running across your environment, govern who can use what, enforce guardrails that prevent harm, and stop risky AI behavior in real time, without slowing your teams down.

By the end of this guide, you'll be able to:

- Explain the AI Security Gap and why it lives on the endpoint
- Tell shadow AI apart from compromised AI, and why both bypass legacy controls
- See why network, CASB, and browser tools structurally miss on-device AI
- Recognize real AI-driven attacks and how to stop them
- Apply a four-part model – Discover, Govern, Apply Guardrails, React
- Build a practical rollout plan for endpoint-native AI governance

Contents

Chapter 1
The AI Blind Spot on Every Endpoint

In just two years, AI has gone from novelty to infrastructure. An estimated 51% of enterprises already run AI agents in production, and 85% will have deployed or planned them by the end of 2026. 40% of enterprise applications will embed AI agents, up from 5% in 2025.
Productivity is soaring. So is risk.

The problem is where AI runs. Employees, enterprise AI agents, and AI-wielding attackers all converge on the endpoint, and none of them are governed by legacy controls. The tools built to watch people and devices simply weren't designed to govern what AI can reach, execute, or weaponize on the device itself.

Firewalls, identity, and EDR watch people and devices. None of them govern what AI can reach, execute, or weaponize on the endpoint.

Security teams call the result the AI Security Gap; the growing inability to govern AI activity, whether it comes from an external attacker or an employee's own tools. It rests on three structural failures: a visibility gap (no reliable inventory of AI on endpoints), a control gap (no runtime enforcement of acceptable AI behavior), and a prevention gap (detection is reactive, while AI acts at machine speed). The chapters ahead unpack each, and how to close them.

Chapter 2

Two Faces of AI Risk: Shadow AI & Compromised AI

AI risk shows up in two distinct ways. Both bypass legacy controls. Both act on the endpoint.

Shadow AI – Unmanaged AI, Outside IT View

Staff paste confidential data into AI assistants and run AI desktop apps and local LLMs outside IT oversight, touching files and workflows directly on the device. Developers connect AI coding tools and MCP servers without governance. None of it shows up in a central inventory, and none of it is bound by policy.

Compromised AI – Trusted AI, then Hijacked

Enterprise AI agents run with broad, standing permissions. A poisoned prompt, a manipulated tool call, or a supply-chain compromise inherits that trust – and the malicious commands look completely legitimate. Because it happens inside an approved app and workflow, it rarely trips a traditional alert.

The feedback loop to watch

- The more AI you adopt, the larger your attack surface becomes.
- The more attackers leverage AI, the more scalable and sophisticated attacks become.
- Without endpoint-level governance, that loop accelerates risk exponentially.

Chapter 3

Why Your Current Tools Can't Govern AI

It's tempting to extend the tools you already own (network gateways, CASB, and enterprise-browser controls), to cover AI. They help with sanctioned, routed SaaS traffic, but they share one blind spot: they only see what passes through them. AI increasingly does not.

- **Local LLMs and CLI agents** – Ollama and command-line AI run entirely on the device, never crossing a gateway.
- **IDE-embedded and desktop AI** – Copilot, Cursor, and desktop assistants act inside trusted apps – invisible to a browser extension.
- **MCP connectors** – Agent-to-tool connections grant file- and system-level reach that traffic inspection never sees.
- **Encrypted and air-gapped activity** – Encrypted API calls hide content; offline endpoints are beyond network reach entirely.

Endpoint detection-and-response tools see processes generically. They can't tell Cursor from a shell, have no per-tool baseline, and offer no AI-aware guardrails. The honest summary: network tools are blind to local execution, and detection tools mostly classify and react after the fact. Neither governs AI at the moment it acts.

Capability	Network / CASB / Browser AI Tools	Endpoint-Native AI Governance Tools
Visibility	Routed traffic or browser tab only	Local LLMs, CLI agents, IDE & desktop AI, MCP
Enforcement point	After traffic leaves the device	At execution, on the endpoint
Action on risk	Detect, classify, log	Default-deny prevention – stops the action
New agent required	Often a proxy, gateway, or browser	None – rides the agent already deployed
Offline / air-gapped	Blind	Fully covered

Chapter 4
AI Attacks Already in the Wild

These aren't hypotheticals. Across 2025, researchers documented AI weaponized at every stage of the attack chain. Here are four case studies, outlining how endpoint-native governance stops each.

PromptLock – polymorphic AI ransomware

Documented by ESET in August 2025, PromptLock connects to a local Ollama endpoint and generates polymorphic Lua at runtime – unique code every run, defeating static EDR signatures. Endpoint-native governance auto-flags Ollama as an AI agent; a file-write spike plus child-process spawn rate triggers a Critical alert, and the local LLM is blocked by policy.

QUIETVAULT – AI-weaponized supply-chain theft

Reported by Google Threat Intelligence, QUIETVAULT arrived via a compromised npm plugin and weaponized pre-installed AI CLIs on developer workstations to recursively scan and exfiltrate GitHub tokens and crypto wallets. Day-1 guardrails block SSH-key and credential-directory access – no ML baseline required.

GTG-2002 – autonomous extortion

In a campaign affecting 17 confirmed organizations, a threat actor ran Claude Code on an attacker-controlled host; the agent autonomously selected exfiltration targets, analyzed financial data, and set six-figure ransoms. Governance identifies Claude Code instantly; a data-volume and sensitive-directory anomaly raises a High alert, and egress policy blocks unapproved LLM services.

Unconstrained Deletion – a runaway agent, no malware required

Told to clean up old records, a coding agent interpreted the task at maximum scope and deleted the entire production database and all backups in under ten seconds. Governance detects the MCP connector's DELETE access pre-emptively, blocks AI from backup locations, and flags mass deletion as critical, stopping the action before it completes.

The common thread: the damage happened on the endpoint, at machine speed, inside trusted tools. Only a control that stops AI at execution can intervene in time.

Chapter 5

A New Model: Discover, Govern, Apply Guardrails, React

Closing the AI Security Gap means moving security to the point of execution and pairing visibility with enforcement. Endpoint-native AI governance does this through four continuous moves.

- **Discover** – Continuously inventory every AI tool, account, agent, browser extension, LLM service, and MCP connector, including shadow AI.
- **Govern** – Map every AI action to user identity, AI service account, and device context; decide which AI may operate, by role or department.
- **Apply Guardrails** – Apply least-privilege policies that block risky actions at runtime: credentials and PII, backup and recovery writes, and risky spawns like PsExec, vssadmin, and PowerShell.
- **React** – Give each AI tool its own local behavioral baseline and flag anomalies in real time; file-op spikes, exfiltration, abnormal connector usage, suspicious execution chains.

Notice what this is not: it is not prompt-reading or data-stream decryption. Governance focuses on what the AI does, not what it says, which keeps it resilient to encryption, payload variability, and the legitimate-looking nature of AI activity, while sidestepping the privacy concerns of content inspection.

51%
of enterprises run AI agents in production today

40%
of enterprise apps will integrate AI agents by end of 2026

44%
of all data breaches now involve ransomware

Sources: IDC / industry survey 2026; Gartner, Aug 2025; Verizon DBIR 2025.

Chapter 6

What Endpoint-Native AI Governance Looks Like

Morphisec AI Usage Control (AIUC) puts this model into practice. It's embedded in the Morphisec Protector across Windows, Linux, and macOS (no new agent, no proxy, no cloud relay), so it governs the CLI agents, local LLMs, IDE plugins, and desktop AI that other tools miss, including offline and air-gapped endpoints. AIUC extends the Anti-Ransomware Assurance Suite, powered by Automated Moving Target Defense (AMTD).

What you gain

- **Full AI Inventory** – Every tool, user, and MCP connector – including everything IT never approved.
- **Ransomware Resilience** – Block pre-encryption AI recon, backup deletion, LoLBin abuse, and credential theft before damage.
- **Compliance Evidence** – Audit events and inventory exports ready for EU AI Act, NIST AI RMF, ISO 42001, and SOC 2.
- **Safe AI Enablement** – Engineering keeps Copilot, Claude, and Cursor; guardrails target risky actions, not the tools.

AI coverage at a glance

AIUC discovers and governs AI across the categories employees and attackers actually use:

- **General assistants** – Microsoft Copilot, ChatGPT, Claude, Gemini, Grok, DeepSeek.
- **Code AI** – GitHub Copilot, Cursor, Codex, Amazon Q, Cody, Windsurf, Cline, Aider, Tabnine, Replit.
- **Local LLM platforms** – Ollama, LM Studio, AnythingLLM, GPT4All, Jan.ai, OpenCode.
- **Enterprise & workspace** – Teams/Office, Notion AI, Glean, Salesforce Einstein, Databricks AI.
- **Malicious AI** – WormGPT, FraudGPT, GhostGPT, LameHug, PROMPTFLUX, Evil-GPT and other adversarial tooling.

Chapter 7
Your AI Governance Rollout Plan

Endpoint-native governance is an expansion of protection you may already run, not a rip-and-replace. A practical sequence:

A five-step rollout checklist

- **Turn on Discover first** – inventory AI across endpoints where anti-ransomware is already enabled – start with shadow AI.
- **Tie AI actions to identity** – map each action to a user, service account, and device, and set policy by role or department.
- **Add least-privilege guardrails** – block AI access to credentials, PII, and backups, and stop risky process spawns.
- **Baseline behavior** – give each AI tool a local baseline and alert on exfiltration and anomalies.
- **Operationalize compliance** – export audit-ready evidence for EU AI Act, NIST AI RMF, ISO 42001, and SOC 2.

Handling the common objections

- **"EDR already covers AI."** – EDR sees processes generically – it can't tell Cursor from a shell, has no per-tool baseline, and no AI-aware guardrails.
- **"We use a CASB for shadow AI."** – That's network/SaaS discovery – blind to local LLMs, CLI agents, and IDE-embedded AI. Governance owns the client-side gap.
- **"Why add another agent?"** – There isn't one. Governance rides the Protector you already run.
- **"Our developers will revolt."** – Default-deny is policy-tunable and targets risky actions – not the tools. It's safe enablement, not a ban.

Next Steps

AI is changing the rules – accelerating both productivity and risk, and concentrating that risk on the endpoint where legacy controls can't reach. You don't close the AI Security Gap with another detection layer. You close it by governing AI where it runs: discover what's there, govern who can use it, enforce guardrails that prevent harm, and react to anomalies in real time.

Morphisec AI Usage Control delivers that governance prevention-first and tied directly to anti-ransomware, and on the agent you already run. The fastest place to start is visibility: turn on Discover and see the shadow AI already operating across your endpoints.

About Morphisec

Morphisec is the global leader in prevention-first cyber defense and anti-ransomware protection, delivering preemptive security that stops attacks before execution. Founded in 2014 and headquartered in New York, and powered by Automated Moving Target Defense (AMTD), Morphisec protects millions of endpoints across finance, healthcare, manufacturing, and technology. At Morphisec, we don't just respond – we prevent.

To learn more, visit morphisec.com/demo