WHITE PAPER

CROWDSTRIKE

Securing AI Where It Executes:

The Endpoint Is the New Control Point for AI Agent Security

How CrowdStrike's Falcon Endpoint Security Delivers Runtime Visibility, Governance, and Threat Detection for Agentic AI Systems Including Claude Code, Claude Cowork, and OpenClaw

Table of Contents

Executive Summary

Artificial intelligence has crossed a critical threshold. AI agents are no longer passive tools that respond to human queries. They are autonomous systems that execute terminal commands, modify files, access sensitive data, and trigger complex downstream workflows directly on enterprise endpoints. From developer coding assistants to general-purpose desktop agents, AI now operates with system-level privileges that were once reserved exclusively for human users and administrators.

This shift creates an unprecedented security challenge. Legacy network perimeters, static application controls, and signature-based defenses were never designed to govern autonomous software that behaves indistinguishably from a human operator. The attack surface has expanded from the application layer to the execution layer, and the enforcement point must move with it.

The CrowdStrike Falcon® platform defends the endpoint as the epicenter of AI security. With kernel-level visibility, runtime behavioral analysis, and the industry's deepest telemetry across more than 1,800 distinct AI applications and nearly 160 million unique application instances, CrowdStrike delivers the only platform purpose-built to discover, govern, and protect AI agents at the point where they actually execute. No other vendor combines the depth of endpoint telemetry, the speed of real-time detection, and the breadth of AI-specific controls that CrowdStrike delivers through the Falcon platform.

This white paper examines the security challenges posed by three of the most widely adopted AI agent platforms in 2026 (Anthropic's Claude Code, Anthropic's Claude Cowork, and the open-source OpenClaw framework) and demonstrates how CrowdStrike's existing endpoint detection and response (EDR) technology, extended by CrowdStrike Falcon® AI Detection and Response (AIDR) and our coming product releases, will provide comprehensive runtime protection for each.

The AI Agent Revolution and the New Threat Landscape

The enterprise is undergoing a fundamental transformation in how software operates. AI agents have evolved from conversational interfaces into autonomous execution engines that take real-world actions with real-world consequences. Three platforms exemplify the scope and diversity of this shift.

Claude Code: The Developer's Agentic Coding Partner

Claude Code is Anthropic's agentic coding tool that operates directly within a developer's terminal, integrated development environment (IDE), desktop application, and browser. Unlike traditional code completion tools, Claude Code autonomously reads entire codebases, edits files across multiple repositories, executes shell commands, manages Git workflows, and orchestrates multi-agent task pipelines. The tool leverages the Model Context Protocol (MCP) to connect with external services including Google Drive, Jira, Slack, and custom APIs, and can spawn sub-agents that work on different parts of a task simultaneously.

From a security perspective, Claude Code operates with the full privileges of the developer's user session. It executes arbitrary terminal commands, writes and modifies source code, and can trigger CI/CD pipelines. Security researchers have already identified critical vulnerabilities including CVE-2025-59536, a consent bypass with a CVSS severity rating of 8.7 out of 10, that exploited project-level configuration files to override built-in safety mechanisms.[1] A single malicious commit to a shared repository could compromise every developer who clones the project.

Claude Cowork: The General-Purpose Desktop Productivity Agent

Launched in January 2026, Cowork brings the same agentic architecture that powers Claude Code to non-technical knowledge workers. Cowork runs inside the Claude Desktop application and can read, edit, create, and delete files in any folder the user grants access to. It can sort and rename entire directory structures, extract data from screenshots, draft reports from research materials, and automate repetitive browser-based workflows through the Claude in Chrome connector.

Cowork sessions persist across interactions, meaning the agent maintains context and can be steered remotely from a mobile device while running in the background. This always-on design pattern, where an AI agent continuously operates on a user's machine with filesystem access, creates a new category of endpoint risk that traditional security tools were never designed to monitor. Anthropic itself acknowledges the risk of prompt injection attacks that could alter Cowork's plans through content it encounters, noting that agent safety remains an active area of development.

[1] https://nvd.nist.gov/vuln/detail/CVE-2025-59536

OpenClaw: The Open-Source Personal AI Agent

OpenClaw is the viral open-source personal AI agent that surpassed 100,000 GitHub stars within weeks of its meteoric adoption in early 2026. OpenClaw operates as a local gateway that connects large language models (LLMs) to the user's files, messaging applications, browsers, smart home devices, and productivity tools through more than 100 pre-built skill extensions. Unlike cloud-based AI assistants, OpenClaw runs locally, executes shell commands, manages email, controls browser automation, and accesses enterprise systems, all while operating outside conventional security controls.

The security implications are severe. Security researchers have discovered numerous malicious skills on ClawHub, OpenClaw's community registry, representing the first major supply chain attack targeting AI agent ecosystems. Cisco researchers have warned that community skills can enable silent data exfiltration and prompt injection. One security vendor found that 22% of its customers had employees using OpenClaw on work devices.[2] OpenClaw represents shadow AI at its most consequential: an autonomous agent framework running on corporate endpoints with system-level access, often without security team knowledge.

KEY INSIGHT

These three platforms share a common characteristic: They all execute on the endpoint. Whether it is Claude Code running terminal commands in a developer's session, Cowork modifying files on a knowledge worker's desktop, or OpenClaw automating workflows through shell scripts and browser control, the endpoint is where AI actions happen. The power of AI agents comes from their ability to automate human-driven decision-making and workloads, which is precisely why the endpoint must be the control point for AI security.

[2]https://token.security/blog/the-clawdbot-enterprise-ai-risk-one-in-five-have-it-installed

Why the Endpoint Is the Control Point for AI Security

Legacy security architectures were built around a perimeter model: Firewalls inspect traffic at network boundaries, proxies filter web access, and SIEM systems correlate logs after the fact. These controls assume threats arrive from outside and internal processes are inherently trusted. AI agents invalidate every one of these assumptions.

AI Agents Operate at the Process Level

When Claude Code executes a command, that command runs as a child process of the user's terminal session. When OpenClaw dispatches a skill to manage files or send emails, it spawns local processes with the user's credentials. When Cowork reorganizes a directory, it performs file system operations indistinguishable from manual user activity. Network-layer controls cannot see these actions because they never cross a network boundary. Application-layer controls cannot distinguish them because the AI agent inherits the user's full context and privilege.

Only deep operating system visibility via an endpoint sensor provides you with the capability to observe the actual behavior of AI agents: the specific commands executed, the files accessed, the data written, the network connections initiated, and the child processes spawned. This is the telemetry that CrowdStrike's Falcon sensor has been capturing for over a decade.

Deep OS-level Visibility Is Non-Negotiable

CrowdStrike Falcon captures process trees, file I/O, registry modifications, and network activity in real time. This architectural position provides three critical advantages for AI agent security:

- **Complete process lineage:** Falcon traces the full execution chain from the parent AI application through every child process, script execution, and downstream system call. When Claude Code spawns a sub-agent that triggers a shell command, Falcon captures the entire lineage from the IDE process through the agent runtime to the terminal execution.
- **Behavioral fidelity:** Because Falcon observes behavior directly, it captures what AI agents actually do rather than what they claim to do. Prompt injection attacks that manipulate an agent's behavior are fully visible to Falcon.
- **Real-time enforcement:** Falcon enables immediate response, from alerting security teams to isolating compromised endpoints, without waiting for logs to be collected, forwarded, and correlated. When an AI agent exhibits suspicious behavior, Falcon can contain the threat before it propagates.

The Endpoint Sees What No Other Layer Can

Consider a prompt injection attack against OpenClaw. A malicious skill downloaded from ClawHub contains hidden instructions that cause the agent to exfiltrate sensitive files to an external server. At the network layer, this appears as a standard HTTPS request from a legitimate application. At the application layer, the agent believes it is performing an authorized task. But on the host itself, Falcon observes the suspicious file access pattern, the unexpected network connection, and the process lineage that traces back to the malicious skill and can stop the activity in real time.

This is why the endpoint is not just one of many control points for AI security. It is the control point. There is no better vantage point from which to observe, govern, and protect AI agent behavior than the operating system where those agents execute.

CrowdStrike Falcon: Purpose-Built AI Agent Security

CrowdStrike has spent over a decade building the industry's most comprehensive endpoint telemetry platform. The Falcon sensor already captures the commands, scripts, file activity, and network connections of every application running on the endpoint. With the introduction of EDR AI Runtime Protection, Shadow AI Discovery for Endpoint, and Falcon AIDR, CrowdStrike extends these capabilities specifically to the AI agent threat surface.

EDR AI Runtime Protection

Falcon's EDR AI Runtime Protection delivers runtime visibility of AI behavior at the point of execution. The Falcon sensor captures the behavioral telemetry of all applications running on the endpoint, including agentic applications, and correlates this activity with CrowdStrike's threat intelligence to identify suspicious or malicious behavior in real time.

When suspicious behavior is detected — whether a Claude Code sub-agent is executing an unexpected privileged command, an OpenClaw skill is accessing files outside its expected scope, or a Cowork session is initiating suspicious network connections — human and agentic security teams will be able to trace the activity to the originating process and associated agentic application and act immediately. Response actions include isolating affected endpoints to contain threats before they spread across the environment.

Shadow AI Discovery for Endpoint

CrowdStrike sensors already detect more than 1,800 distinct AI applications running across enterprise devices. Shadow AI Discovery for Endpoint automatically identifies AI applications, agents, LLM runtimes, MCP servers, and development tools running across endpoints and links them to asset context and privilege exposure to prioritize risk to critical systems.

For enterprises where developers may be running Claude Code with unrestricted MCP server connections, or knowledge workers have independently installed Cowork with access to sensitive file shares, or OpenClaw has been deployed on corporate devices without IT approval, Shadow AI Discovery for Endpoint provides the visibility that security teams need to assess not just what AI is deployed but the potential agentic blast radius of a compromise.

Falcon AIDR: AI Detection and Response

Falcon AIDR extends prompt-layer protection to agentic web and desktop AI applications including ChatGPT, Gemini, Claude, DeepSeek, Microsoft Copilot, O365 Copilot, GitHub Copilot, and Cursor. Falcon AIDR delivers real-time prompt inspection, detects prompt attacks and data leaks, and surfaces AI access control violations and AI content violations.

Falcon AIDR maps the relationships between users, prompts, models, agents, and MCP servers to provide comprehensive visibility into how AI systems interact with enterprise data and infrastructure. With attribute-based access controls, security teams can enforce granular AI security policies across users, agents, tools, and models. Falcon AIDR detects and prevents prompt injection attacks with up to 99% efficacy at sub-30 millisecond latency, inspecting every AI system input and output to catch direct and indirect prompt injection attacks and automatically redact sensitive data.

Securing Claude Code, Claude Cowork, and OpenClaw with Falcon

Each of these three AI agent platforms presents distinct security challenges that map directly to CrowdStrike Falcon's capabilities. The following sections detail how Falcon's EDR, Shadow AI Discovery, and AIDR work together to protect enterprises against the specific risks each platform introduces.

Securing Claude Code

Threat Profile

Claude Code executes with full user privileges, creates novel scripts, and runs arbitrary commands. Its MCP integration connects it to external services, and its sub-agent architecture means multiple autonomous processes may be running simultaneously within a single developer session. Known vulnerabilities have demonstrated that malicious repository configurations can hijack the tool's execution before user review.

Falcon Endpoint Capabilities

- **Process tree monitoring:** CrowdStrike Falcon® Insight XDR traces the full execution lineage from the Claude Code parent process through every sub-agent, shell command, and file operation to identify suspicious activity such as unexpected privilege escalation.
- **MCP server discovery:** Shadow AI Discovery identifies all MCP servers connected to Claude Code instances across the enterprise, maps trust relationships and flags ungoverned or unauthorized connections.
- **Prompt injection defense:** Falcon AIDR inspects prompts and responses flowing through Claude Code Desktop, detecting injection attacks that attempt to manipulate the agent's behavior through malicious code comments, configuration files, or repository content.

Securing Claude Cowork

Threat Profile
Claude Cowork operates as a persistent, background agent with direct file system access on non-technical users' devices. It can read, edit, create, and delete files in any folder the user grants access to, and can browse the web through the Claude in Chrome connector. Its persistent session model means it maintains context across interactions and can be remotely steered, creating an always-on attack surface on endpoints that may contain sensitive business data.

Falcon Endpoint Capabilities

- **File system behavioral analysis:** Falcon Insight XDR monitors Cowork's file operations in real time and alerts on suspicious activity such as bulk file access, sensitive data reads, and unexpected file deletion patterns.
- **Browser activity monitoring:** CrowdStrike's acquisition of Seraphic extends runtime protection to the browser layer, securing Cowork's agentic activity when it operates through Claude in Chrome.
- **Data loss prevention:** Falcon AIDR detects and blocks sensitive data, including credentials, regulated information, and proprietary business content, before it can be exposed through Cowork's interactions with AI models or external services.

Securing OpenClaw

Threat Profile
OpenClaw represents the most complex AI agent security challenge in the enterprise. It is open-source, self-hosted, highly extensible, and often deployed without IT oversight. Its 100+ skill ecosystem includes community-contributed extensions that have already been weaponized through supply chain attacks. OpenClaw executes shell commands, manages email, controls browsers, and accesses enterprise systems with user-level privileges. Its agent architecture supports persistent memory, cron-based automation, and multi-agent coordination, all running on endpoints that may also contain sensitive corporate data.

Falcon Endpoint Capabilities

- **Shadow AI detection:** Shadow AI Discovery identifies OpenClaw installations across the enterprise, even when deployed without IT knowledge. Falcon detects the distinctive process signatures, port usage (default 18789), and workspace structures that characterize OpenClaw deployments.
- **File system behavioral analysis:** Falcon Insight XDR monitors OpenClaw's file operations in real time and alerts on suspicious activity such as bulk file access, sensitive data reads, and unexpected file deletion patterns.
- **Automated containment:** When an OpenClaw agent exhibits behavior consistent with prompt injection or skill-based attacks, Falcon Insight XDR can automatically isolate the affected endpoint, preventing lateral movement and data exfiltration while preserving forensic evidence.

Falcon AI Security: Capability Matrix

The following matrix illustrates how CrowdStrike Falcon's capabilities map to the security requirements of each AI agent platform.

Falcon Capability	Claude Code	Cowork	OpenClaw
EDR Runtime Protection	✓	✓	✓
Process Tree Analysis	✓	✓	✓
Shadow AI Discovery	✓	✓	✓
MCP Server Detection	✓	—	✓
AIDR Prompt Inspection	✓	✓	—
Skill/Plugin Monitoring	—	✓	✓
Browser Runtime Protection	NA	✓	✓
Automated Containment	✓	✓	✓
Data Loss Prevention	✓	✓	✓

This capability matrix reflects both existing features as well as functionality and benefits announced at RSA 2026 that are still in development.

Beyond the Endpoint: Falcon for Unified AI Security Across SaaS, Browser, and Cloud

AI agents do not operate exclusively on the endpoint. Claude Code connects to cloud-based CI/CD pipelines and SaaS development platforms. Cowork reaches into the browser through Chrome extensions and interacts with cloud-hosted connectors. OpenClaw integrates with dozens of SaaS applications, messaging platforms, and cloud APIs. CrowdStrike's platform extends AI security across every surface where agents operate.

SaaS Agent Discovery and Governance

CrowdStrike Falcon Shield's SaaS and AI Agent Discovery provides visibility into AI agent activity, configured capabilities, and data access across leading platforms including Microsoft Copilot, Google Gemini, Salesforce Agentforce, ChatGPT Enterprise, ServiceNow, and Box. Agents are attributed to their owners, mapped to the tools and APIs available to them — including invocation of external agents, REST API calls, and MCP server connections — and correlated with who has access to each agent. For organizations where AI agents interact with SaaS-hosted data, this capability closes the visibility gap between endpoint activity and cloud-based operations.

Browser Runtime Protection

CrowdStrike's acquisition of Seraphic extends runtime protection to the browser to secure agentic activity at the point where it increasingly operates. As Cowork and OpenClaw expand their browser automation capabilities, this layer of protection becomes essential for preventing data exfiltration, and unauthorized actions that originate from or traverse browser-based interfaces.

Cloud and Container Security

AIDR secures AI workloads running in containerized environments communicating with the OpenAI API specification, providing runtime inspection for AI services and detection of prompt attacks, data leaks, and policy violations. CrowdStrike Falcon® Cloud Security monitors Claude Code's interactions with build and deployment systems, alerts on unauthorized pipeline triggers, and blocks potentially destructive deployment actions. CrowdStrike Falcon® Data Security for Cloud delivers real-time visibility into how sensitive data flows into and through AI services, enabling teams to identify AI data exposure as it happens and automate response through unified SOAR workflows.

Why CrowdStrike: No Better Company to Secure AI

The AI agent security market is emerging rapidly, and multiple vendors are positioning solutions. CrowdStrike's advantage is structural, not incremental. Three factors distinguish the Falcon platform as the definitive AI security solution for the agentic era.

Unparalleled Endpoint Expertise

CrowdStrike pioneered modern EDR. The Falcon sensor has been capturing behavioral telemetry at the kernel level for over a decade, processing trillions of events daily across millions of endpoints. This investment in depth, not just breadth, means that when a new class of threat emerges (as AI agents represent), CrowdStrike's existing telemetry infrastructure already observes the relevant behavior. No redeployment, no new sensors, no architectural changes. The sensor that protects against adversarial behavior today is the same sensor that protects against AI agent compromise tomorrow.

The Largest AI Threat Telemetry Dataset in the Industry

With visibility into more than 1,800 distinct AI applications and nearly 160 million unique application instances, CrowdStrike possesses the largest dataset of AI agent behavior in the cybersecurity industry. This telemetry powers behavioral models that distinguish between legitimate AI agent activity and indicators of compromise, prompt injection, and data exfiltration. The models improve continuously as the dataset grows, creating a defensibility advantage that compounds over time.

From EDR to AIDR: Architectural Consistency

The Falcon platform delivers a unified security model for AI, protecting everything from the environments where AI runs to the interaction layer where prompts and agents operate. This architectural consistency means that organizations do not need separate tools for endpoint security, AI governance, and prompt-layer protection. One platform, one sensor, one console.

The CrowdStrike Advantage

Security built for static applications cannot keep up with autonomous systems. Organizations need real-time visibility and control over AI behavior wherever it runs. CrowdStrike is that new standard.

— **Michael Sentonas**, President, CrowdStrike

Recommendations for Enterprise Security Leaders

As AI agents proliferate across enterprise endpoints, security leaders should take immediate action to establish visibility and governance. CrowdStrike recommends the following framework for securing AI agent deployments.

1. **Discover your AI footprint.** Deploy Shadow AI Discovery to identify every AI application, agent, LLM runtime, MCP server, and development tool running across your endpoints. You cannot secure what you cannot see.

2. **Leverage endpoint security.** Use EDR to gain visibility into agentic activity, detect behavior indicative of attacks, and respond in real time.

3. **Deploy prompt-layer protection.** Activate Falcon AIDR to inspect prompts and responses in real time, detecting injection attacks, data leaks, and policy violations across Claude Code, Cowork, and other agentic AI applications on the endpoint.

4. **Govern AI agent permissions and data access.** Map the blast radius of each AI agent deployment by linking agent discovery to asset context and privilege exposure. Prioritize remediation for agents with access to critical systems and sensitive data.

5. **Extend protection beyond the endpoint.** Enable SaaS, browser, and cloud AI security capabilities to govern agent behavior across every surface where AI operates, ensuring consistent policy enforcement from endpoint to cloud.

Conclusion

The rise of AI agents represents the most significant shift in enterprise computing since the adoption of cloud infrastructure. Claude Code is redefining how software is built. Claude Cowork is transforming how knowledge work is performed. OpenClaw is demonstrating how open-source agents can automate virtually any workflow. Each of these platforms creates extraordinary productivity gains, and each introduces security risks that legacy tools cannot address.

The common thread is the endpoint. AI agents execute on devices. They run commands, access files, initiate network connections, and interact with sensitive data at the operating system level. This is where security must be enforced, not at the network perimeter, not at the application layer, but at the operating system level where behavior is observable, attributable, and controllable in real time.

CrowdStrike Falcon is purpose-built for this moment. With deep OS-level visibility via EDR, AI-specific runtime protection, comprehensive shadow AI discovery, and prompt-layer detection and response through AIDR, CrowdStrike delivers the only platform that secures AI agents at the point of execution and extends that protection across SaaS, browser, and cloud environments.

The endpoint is the epicenter of agentic security. CrowdStrike is the company that secures it.
To learn more, visit https://www.crowdstrike.com/en-us/solutions/secure-your-ai/

Disclaimer This white paper includes discussion of unreleased services and features. Any references to unreleased features reflect our current plans only and do not constitute a promise or commitment to deliver such features. These items may change or may not be made available in all regions. Customers should make purchase decisions based on features currently available.

About CrowdStrike

CrowdStrike (NASDAQ: CRWD), a global cybersecurity leader, has redefined modern security with the world's most advanced cloud-native platform for protecting critical areas of enterprise risk — endpoints and cloud workloads, identity and data.

Powered by the CrowdStrike Security Cloud and world-class AI, the CrowdStrike Falcon platform leverages real-time indicators of attack, threat intelligence, evolving adversary tradecraft, and enriched telemetry from across the enterprise to deliver hyper-accurate detections, automated protection and remediation, elite threat hunting, and prioritized observability of vulnerabilities.

Purpose-built in the cloud with a single lightweight-agent architecture, the Falcon platform delivers rapid and scalable deployment, superior protection and performance, reduced complexity and immediate time-to-value.

CrowdStrike: We stop breaches.

Learn more: https://www.crowdstrike.com/en-us/

Start a free trial: https://www.crowdstrike.com/free-trial-guide/

Follow us: Blog | X | LinkedIn | Facebook | Instagram