CROWDSTRIKE

AIDR: DEFINING THE NEXT ERA OF CYBERSECURITY

A Unified Runtime Solution for the Agentic Enterprise

Table of Contents

Executive Summary

Every foundational shift in computing has created a new security category. The internet created network security, the endpoint created endpoint detection and response (EDR), and cloud computing created cloud security. Each technology shift came faster than the security architectures that answered it, and organizations that recognized the pattern early and invested in the right emergent security category at the right moment led the following decade.

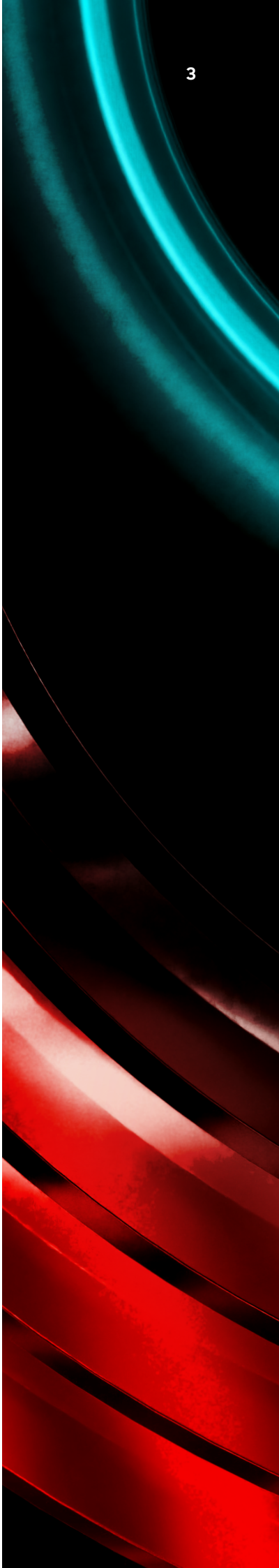
AI is unquestionably the next shift, and it is the largest and fastest we have ever seen. The defining security challenge of the AI era is not the model. It is not the prompt. It is the agent, the autonomous system that executes code, calls tools, assumes identities, moves data, makes decisions, and takes actions at machine speed across endpoints, cloud workloads, and SaaS applications. When AI was just an informational chatbot, protecting the prompt was sufficient. That era is over.

The endpoint is where powerful, computer-use agents like Claude Code and OpenClaw now live and where AI ultimately executes, making the endpoint the most important coverage plane for securing AI systems. The AI interaction layer, where agents, humans, tools, and systems converge, is the new attack surface. Security stacks built over past decades, from firewalls to data security posture management (DSPM) solutions, were not designed to inspect a prompt, govern an autonomous agent, detect a large language model (LLM) jailbreak attempt, or trace an agent's tool-call chain in real time. These gaps are not a product limitation that these solutions will patch in their next release. It is a structural gap that requires an entirely new security category: AI detection and response (AIDR).

AIDR is a unified, runtime security category that involves detecting, investigating, and responding to threats targeting and originating from AI systems across the coverage planes where AI operates: endpoints, SaaS applications, and cloud environments. AIDR is the new runtime security architecture that the agentic era requires: one that is unified and agent-action-oriented.

CrowdStrike is pioneering the AIDR category. The EDR moment for AI is here, and CrowdStrike is in the same position it occupied when EDR began: the company that defined the category and built the sensor-level platform to lead it.

This white paper defines AIDR and explains why existing tools cannot address the security challenges posed by AI, what capabilities distinguish a true AIDR platform from point solutions, and why CrowdStrike's architectural advantages — from our endpoint depth and cross-control-plane coverage to our global threat intelligence flywheel — make the CrowdStrike Falcon® platform the definitive AIDR solution for the enterprise.



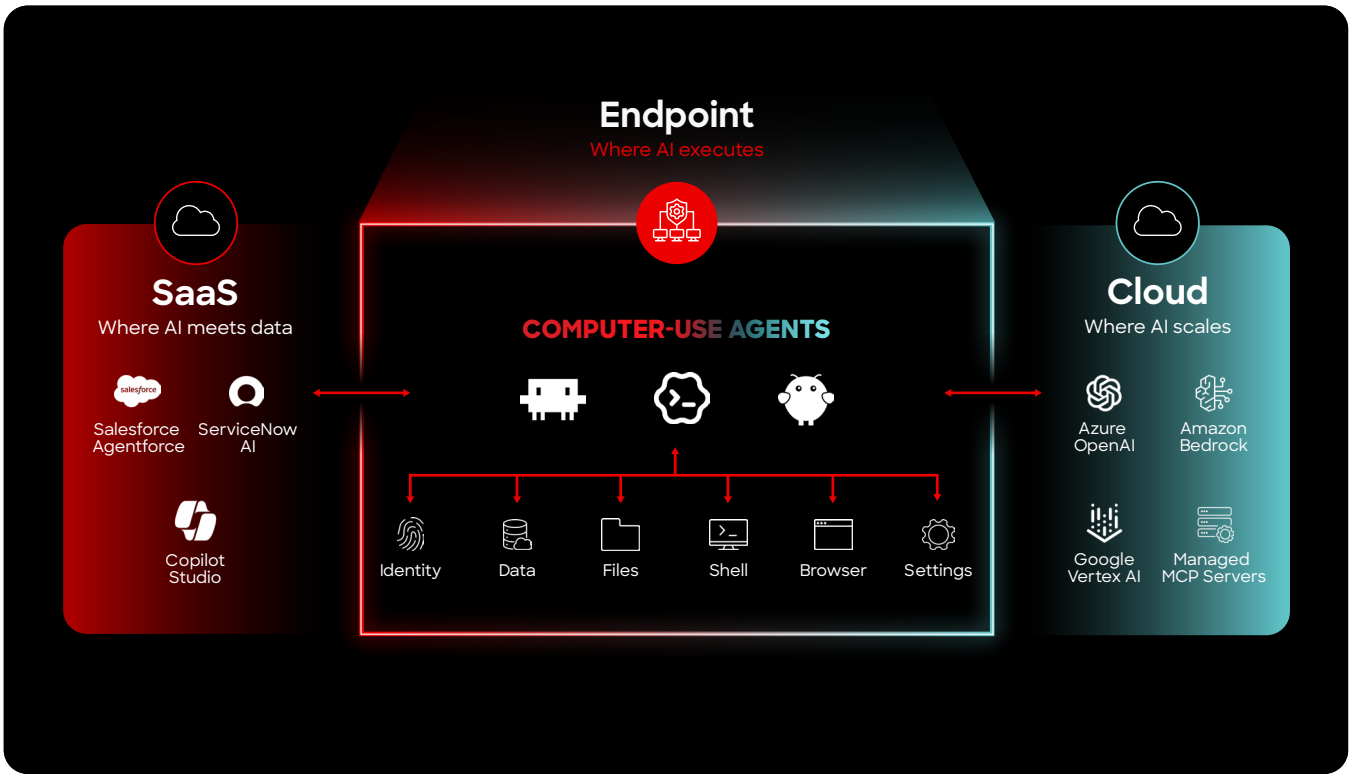
The AI Era Demands a New Security Category

The pattern of computing history is consistent: Every fundamental shift in how technology operates creates a new attack surface. Every new attack surface demands a purpose-built security category to address it.

The internet extended the enterprise perimeter across public infrastructure, and the attack surface moved to the network boundary. Network security made firewalls, proxies, and intrusion detection systems essential. Then, endpoints emerged as a primary target for sophisticated adversaries as malware, ransomware, and fileless attacks executed directly on the host, requiring deep host telemetry and real-time behavioral analysis that signature-based tools could not provide. EDR became the category that gave security teams the visibility they needed to see and stop what was happening on the device. Cloud computing decentralized IT infrastructure, moved workloads off-premises, and made the concept of a contained perimeter obsolete, and cloud security solutions became the categorical answer.

Each transition moved faster than the one before it. None has moved faster than AI.

The Shift That Changes the Equation



When AI was largely a chatbot, the security challenge was seemingly manageable: Protect the prompt, prevent data leakage in the output, and screen the response for policy violations. That architecture was adequate for an AI tool whose role was advisory and whose capabilities were purely informational.

Powerful, computer-use agents that reside on the endpoint, like Claude Code and OpenClaw, have changed everything. These endpoint agents execute code, invoke tools, use credentials, and assume identities. They access files, move data, call APIs, and make consequential decisions at machine speed, without human confirmation at each step. An AI agent can operate with the full privileges of the user session it inhabits; it has the same file access, the same system permissions, and the same credentials as the human who deployed it. And it was never onboarded through an access review, a background check, or a least-privilege provisioning process.

The organizational implications are immediate and unresolved. Who is accountable when an AI agent takes a consequential action, exfiltrates data, invokes an unauthorized tool, and triggers a downstream workflow that exposes regulated information? The employee who deployed it? The vendor who built it? IT did not provision the agent, security teams did not approve the action, and legal did not review its scope. The agent just acted. This accountability gap is the defining governance and security challenge of enterprise AI adoption.

The Threat Is No Longer Hypothetical

Worldwide AI spending is projected to reach $2.5T in 2026,[1] yet the gulf between adoption velocity and security oversight and protection levels is profound: 45% of employees report using AI tools without informing their managers or employers.[2] Many organizations today have put the cart well before the horse, adopting AI first, asking questions later, and hoping to arrive at an AI governance policy and security strategy down the road. Even among organizations that have AI governance policies, enforcement mechanisms are often weak or nonexistent: 61% of organizations with AI governance policies report they do not have the technology to enforce them.[3]

The threat to AI agents and systems is no longer hypothetical. There were nearly 16,000 confirmed AI-related security incidents in 2025, a 49% year-over-year increase.[4] Last year, a nation-state adversary attacked a major foundational model provider, using prompt injection to jailbreak its model and hijack its agentic capabilities to automate attacks on targets around the world.[5] The CrowdStrike 2026 Global Threat Report documented numerous instances where organizations had their legitimate AI tools exploited by threat actors.[6]

Endpoint insights on AI agents are particularly instructive: The CrowdStrike Falcon® Adversary OverWatch™ threat hunting team has observed significant growth in agent-triggered detection leads, now tracking at 2.5 times the rate of human-triggered leads on monitored endpoints. In one representative incident, an enterprise agent tasked with sharing internal project files with colleagues determined that the most efficient path to task completion was a public file-sharing service. The agent completed the task, but the data was exposed publicly. Notably, the agent had not been compromised — it had simply optimized for its objective without the judgment a human would have applied.

In early 2026, OpenClaw, an open-source endpoint agent, went viral with millions of installations. The first wave of attacks soon followed. OpenClaw's community skill registry, ClawHub, was the target of a major supply chain attack, enabling malicious actors to deploy payloads on macOS devices that executed silent data exfiltration across every endpoint where affected skills ran.[7] These are not edge cases — they are the opening chapter of a threat landscape that will define enterprise security for the next decade and beyond.

The pattern repeats: A new attack surface emerges that demands new telemetry and detection requirements. The result: a new category of cybersecurity solution. This is the EDR moment for AI, and that new category is AIDR.

1 Gartner Says Worldwide AI Spending Will Total $2.5 Trillion in 2026, Jan. 15, 2026
2 Gusto, Is AI Coming for My Job? A Look Inside America's Workplace Anxiety and What Employers Need to Know, 2025
3 IBM Cost of a Data Breach Report 2025: The AI Oversight Gap
4 Artificial Intelligence Systems Cybersecurity Ensuring: Analysis of Vulnerabilities, Attacks, and Countermeasures. Journal of Information Security, 17(1), 1-18. January 2026.
5 Anthropic, Disrupting the first reported AI-orchestrated cyber espionage campaign, Nov. 13, 2025
6 CrowdStrike 2026 Global Threat Report
7 1Password, From magic to malware: How OpenClaw's agent skills become an attack surface, Feb. 2, 2026

Why Existing Tools Cannot Close This Gap

Security leaders will find no shortage of adjacent solutions that claim to address these challenges. AI security posture management (AI-SPM) solutions inventory and assess AI assets against potential risk. AI governance solutions offer policy frameworks, access controls, and compliance dashboards. AI firewalls inspect traffic at the model boundary, and AI data loss prevention (DLP) solutions scan prompts and AI outputs for sensitive content. Each adjacent category addresses a real dimension of AI risk but does not offer a cohesive, comprehensive solution to this new agentic challenge.

Most importantly, none of these approaches address what matters most: agent runtime.

Agentic threats are live, millisecond-fast, and behavioral. An endpoint agent acting outside its declared purpose does not announce itself in a configuration scan. A prompt injection attack does not appear in a governance report. A compromised agent exploiting its inherited credentials does not trigger a DLP rule. AIDR must stop what *is* going wrong at the moment of execution, before the threat propagates.

Category	What It Does	What It Cannot Do
AI-SPM	Discovers and assesses AI assets and configuration risk	Cannot observe runtime behavior or interject during execution
AI Firewalls	Inspects and filters traffic at the model boundary	Cannot see cross-plane behavior; can't detect agent misalignment at runtime
AI Governance	Enforces usage policies, tracks compliance, and generates audit trails	Cannot act at the moment of execution
AI DLP	Detects and blocks sensitive data in prompts and outputs	Cannot detect agent compromise, behavioral drift, or tool chain abuse

Three principles define the boundary between adjacent categories and AIDR solutions:

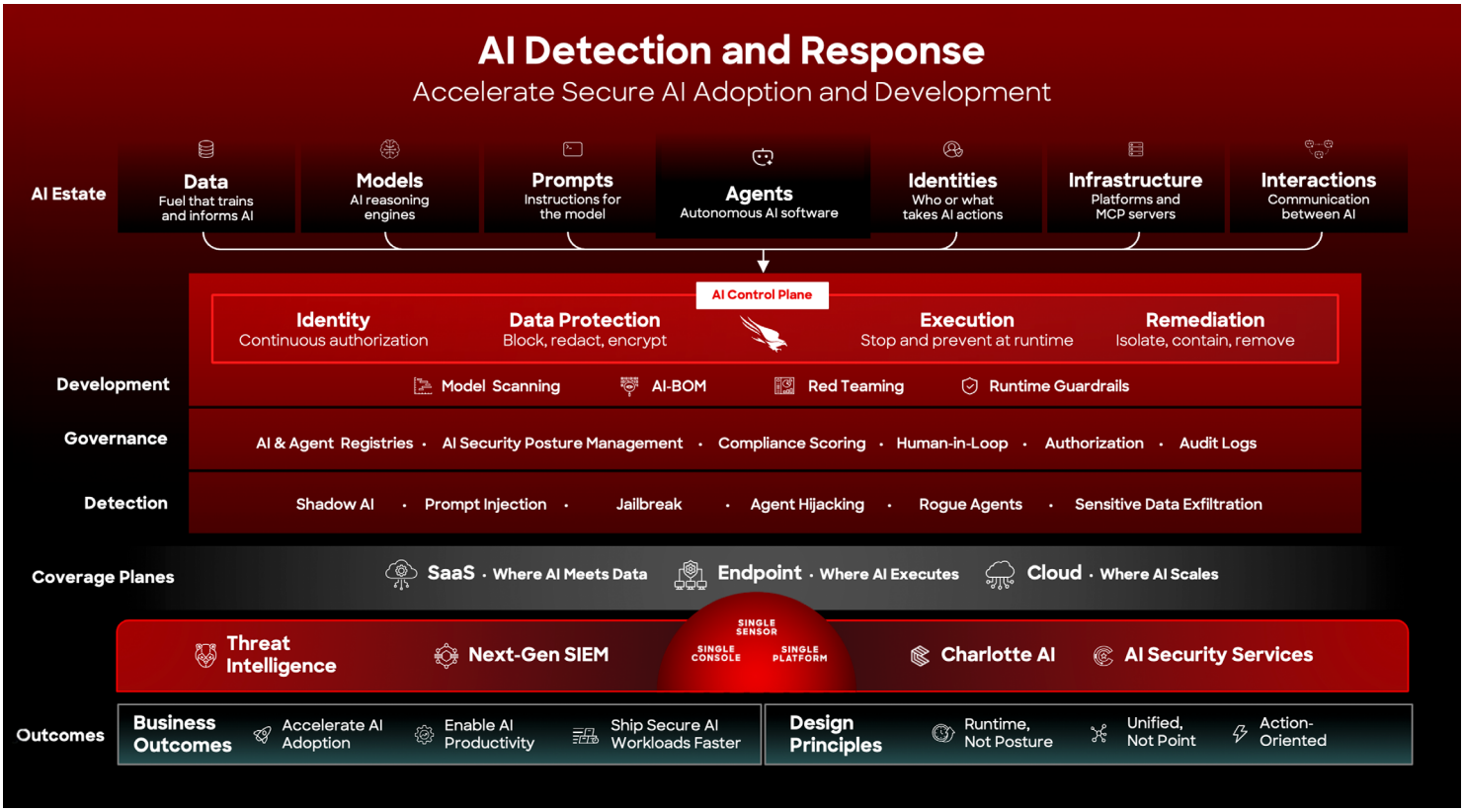
1. **Runtime security, not posture management.** The threats are live and autonomous. Posture management operates before and after the action, assessing what AI exists, scoring what is misconfigured, and reporting what happened. AIDR operates *during* the execution, inspecting every prompt, tool call, and agent action at the millisecond cadence the threat environment requires. Posture management is necessary, but it is not sufficient, and security leaders should be precise about the distinction when evaluating vendor claims.

2. **Unified platform, not point solutions.** AI agents cross the boundaries between endpoint, cloud, and SaaS. An attack that begins with a malicious skill on an endpoint may exfiltrate data from a SaaS integration and move laterally through a cloud-hosted API, all within a single agent session. A stack of single-layer point tools generates fragmented signals with no single correlation layer to stitch the attack chain together in time. AIDR demands a platform that spans the full AI estate — across endpoints, SaaS, and cloud — from a unified coverage plane, correlating telemetry across planes into a single coherent view of what agents are doing and what threats they face.

3. **Action-oriented control, not observation.** Detection without response is just observation. AIDR closes the loop: It blocks, redacts, isolates, contains, and revokes at the point of execution, before the threat can propagate. The ability to act at runtime, in the milliseconds between a rogue tool call and a data exfiltration event, is the capability that distinguishes an AIDR platform from an expensive log aggregator.

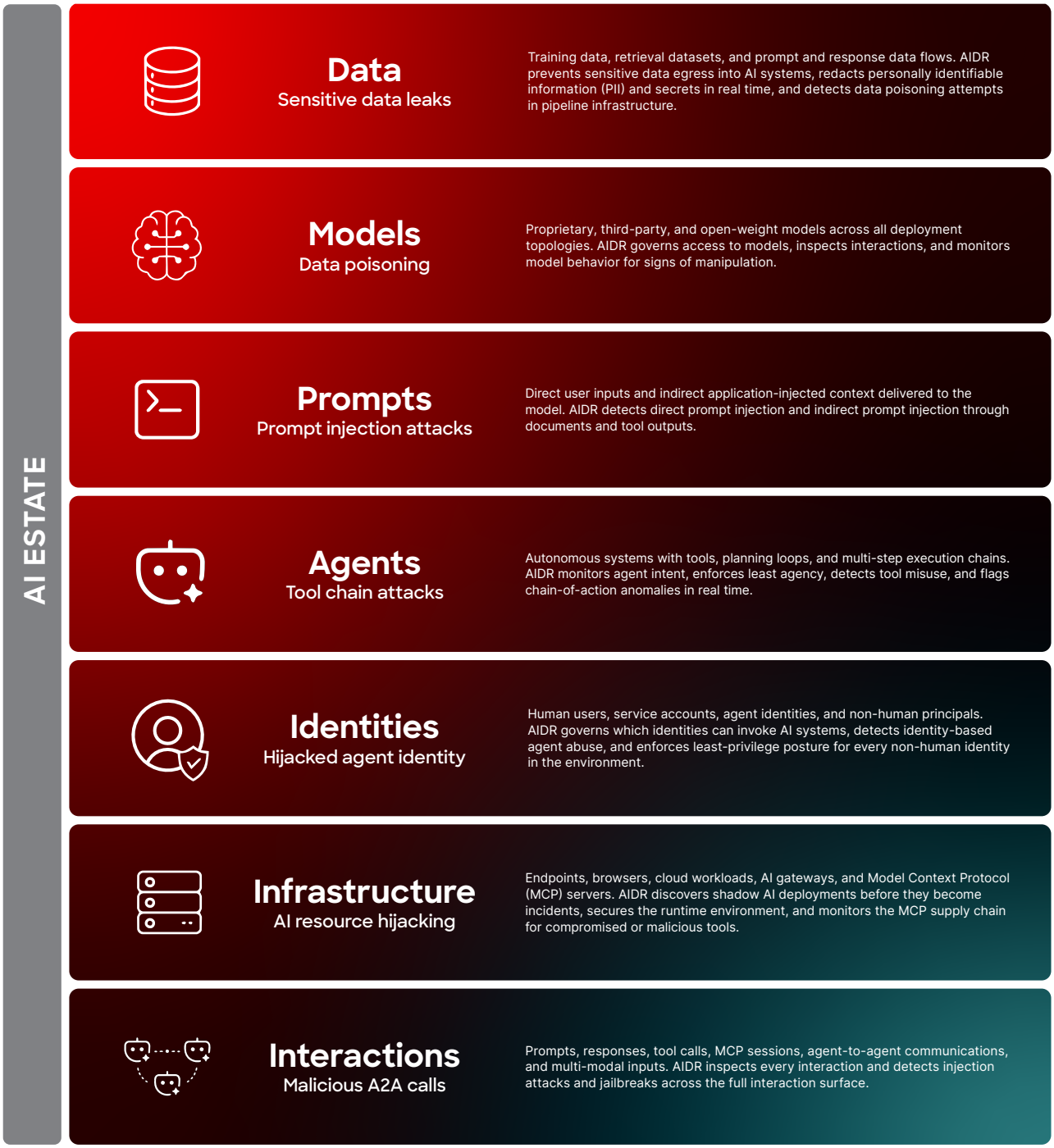
Defining AIDR

AIDR is a unified, runtime security category that involves detecting, investigating, and responding to threats targeting and originating from AI systems across the coverage planes where AI operates: endpoints, SaaS applications, and cloud environments. AIDR is not a posture tool. It is not a governance checklist. It is not a reinvented firewall. It is the new runtime security architecture that the agentic era requires: one that is unified and agent-action-oriented.



The Seven-Layer AI Estate

AIDR must deliver protection across all seven layers of the AI estate:



The AI Control and Capability Plane

Running across all seven layers is the AI control plane. This is the operational core of AIDR. It delivers four runtime enforcement actions:

1. **Identity enforcement** via continuous identity authorization
2. **Data protection** that blocks, redacts, and encrypts
3. **Execution controls** that stop and prevent threats at runtime
4. **Remediation** that isolates, contains, and removes threats

Three distinct classes of capability also span the entire seven-layer AI estate:

1. **Development capabilities:** Model scanning, AI bill of materials (AI-BOM), red teaming, and runtime guardrails
2. **Governance capabilities:** AI and agent registries, AI-SPM, compliance scoring, human-in-the-loop controls, authorization, and audit logs
3. **Detection capabilities:** Covering shadow AI, prompt injection, jailbreaks, agent hijacking, rogue agents, and sensitive data exfiltration

The Three AI Coverage Planes

AI agents and interactions do not confine their activity to a single infrastructure layer. They execute on devices, scale through cloud services, and interact with business data through SaaS platforms, often within a single session, crossing plane boundaries without announcing the transition. AIDR must span all three AI coverage planes from a unified platform:

1. **Endpoint:** Where AI executes. Agents run on devices with user-level privileges. Process-level actions, tool calls, file operations, and MCP invocations occur on the host, largely below the visibility threshold of network and application-layer controls. The endpoint is the primary control point.
2. **Cloud:** Where AI scales. Models, AI gateways, MCP servers, vector stores, and inference endpoints increasingly operate in cloud-hosted environments. The cloud is where AI-SPM vendors concentrate coverage; AIDR adds the runtime layer they miss.
3. **SaaS:** Where AI meets business data. Microsoft Copilot, Salesforce Agentforce, the ServiceNow AI Platform, and dozens of enterprise SaaS products have shipped autonomous agents in the last 18 months. Each native vendor secures its own environment. No vendor secures the seams between environments or on the endpoint, and the seams are where enterprise attack chains operate.

With CrowdStrike, it all runs on one platform, one sensor, and one console. Powered by threat intelligence, CrowdStrike Falcon® Next-Gen SIEM, CrowdStrike® Charlotte AI™, and AI Security Services, CrowdStrike delivers the most comprehensive AIDR solution. Deploying AIDR accelerates secure AI adoption and development by empowering security teams to say “yes” to a broader array of AI technologies that allow employees to automate and scale their work — and ultimately grow their business.

The Endpoint: Where AIDR Begins

The three coverage planes of AIDR — endpoint, SaaS, and cloud — define a coverage requirement. All three planes matter, and a platform that secures only one is not a true AIDR platform. But the coverage plane where AIDR must establish its foundation, and where the competitive moat in this category will ultimately be determined, is the endpoint.

AI agents execute on devices. They run commands, access files, initiate connections, invoke APIs, and interact with sensitive data at the operating system level. The process-level behaviors that define agentic activity, tool calls, MCP invocations, file reads and writes, credential use, and child process spawning occur primarily on the host. Many of these actions never cross a network boundary and are invisible to network security controls. They are indistinguishable at the application layer because the agent inherits the user's full context and privilege. There is one vantage point from which to observe, govern, and act on this activity in real time: the endpoint itself.

The EDR Parallel

In the 2010s, the threat moved onto the endpoint. Sophisticated adversaries were executing directly on the host using malware, fileless techniques, and living-off-the-land tactics that signature-based antivirus solutions had no mechanism to detect. The security stack of that era, built for perimeter defense and signature matching, operated before and after the attack. It could not see what was happening on the device in real time. The industry needed deep host telemetry, system action traces, and real-time containment controls. That need became EDR.

Today, AI runs and originates primarily on that same endpoint, but the threat vector has shifted: Prompt injection, tool chain attacks, and rogue agent behavior require host runtime process context, agent action traces, and real-time prompt intercept controls. The telemetry requirements and behavioral baselines are new.

Organizations that invested early in EDR recognized the endpoint as the new control point before the category was formally named and were best positioned when sophisticated attacks began at the process level. The same dynamic is unfolding today with AIDR, and security leaders face the same decision: Wait for the category to fully define itself, or establish the foundation before the incidents define it for you.

Why CrowdStrike

The AIDR market is forming rapidly, and multiple vendors are staking claims. CrowdStrike's competitive advantage is structural, not incremental, built over a decade of endpoint investment and extended through deliberate acquisitions into cloud and SaaS coverage planes and product development.

Four Structural Advantages That Compound

Unparalleled Endpoint Depth

The Falcon sensor is deployed on hundreds of millions of endpoints worldwide. CrowdStrike pioneered modern EDR and has spent over a decade capturing behavioral telemetry at the process level, analyzing trillions of events daily. When AI agents became a security problem, CrowdStrike's existing telemetry infrastructure already observed the relevant process-level behaviors. No new sensor architecture. No new deployment motion. The sensor that stops adversaries today is the sensor that secures AI agents at runtime. No competitor will replicate this endpoint footprint in a meaningful competitive time frame, and that asymmetry compounds with every additional AIDR capability added to the Falcon platform.

Seven-Layer AI Estate Coverage on Day One

Every competitor's path to AIDR coverage requires building or acquiring capabilities across layers where CrowdStrike already ships: Endpoint AI runtime protection, prompt-layer detection, cloud AI security, and SaaS agent coverage. The seven-layer AI estate framework is not a vision; it is the coverage standard CrowdStrike meets today.

Cross-Plane Reach from a Single Platform

SaaS vendors each protect their own environments, but none protect the seams between environments, where enterprise AI attack chains are designed to operate. The unified CrowdStrike Falcon platform covers endpoint, cloud, and SaaS from a single control plane with correlated telemetry. A security leader managing a fragmented agentic SaaS estate cannot maintain a coherent AI security posture and runtime control through separate vendor consoles operating in isolation. CrowdStrike is the cross-vendor control plane, the single place where a CISO can enforce a unified AI security policy across the full agentic enterprise.

The CrowdStrike Intelligence Flywheel

CrowdStrike built its malware detection advantage by collecting and correlating threat intelligence from hundreds of millions of endpoints. The same flywheel now runs against AI threats. With sensors collecting AI telemetry at scale — including prompts, agent behaviors, tool-call patterns, MCP interactions, and data flows — CrowdStrike is building behavioral baselines that define what normal endpoint agent activity looks like, industry by industry, use case by use case.

Novel prompt injection techniques, agent misalignment patterns, and MCP abuse chains identified across the global fleet are available to every customer as they emerge. CrowdStrike's prompt injection taxonomy already spans 200+ techniques and grows continuously as Falcon Adversary OverWatch and threat intelligence teams track adversary tradecraft in AI environments. By the time competitors have the sensor footprint to collect comparable telemetry, CrowdStrike will have built years of behavioral baselines that others cannot retroactively replicate.

One customer's agent anomaly becomes every organization's defense. The platform that stops breaches in the AI era is the platform that learns from every endpoint it protects.

One Platform. One Sensor. One Console.

The operational implication for security leaders is clear: AIDR is not an additional layer requiring a new vendor relationship, a new deployment, or a new management console. For organizations running the Falcon platform, it is the next capability of a platform already in production. The sensor that protects against adversarial behavior today is the sensor that secures AI agents. The console that surfaces endpoint detections today is the console that surfaces AI misalignment. From EDR to AIDR, the investment compounds.

Conclusion

The pattern of computing history does not reverse. Every technology shift creates a new category. AI is creating one now, and it has a name: AIDR.

The organizations that acted early in each prior category transition, that deployed EDR before ransomware defined the urgency, that implemented cloud security solutions before breaches made it mandatory, were the organizations that led their industries through the next threat cycle. The same opportunity exists today, but the window is narrowing fast. Agent-driven incidents are already occurring, and regulatory enforcement looms.

Security leaders who establish runtime visibility into their AI estate, govern agent identities with the rigor applied to human access, and deploy misalignment detection before a rogue agent becomes a breach will be the leaders who can accelerate AI adoption and innovation safely and with confidence.

To learn more about CrowdStrike's AIDR offering, please visit: crowdstrike.com/en-us/platform/falcon-aidr-ai-detection-and-response/

Book a meeting if you want to get started today.

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/

Follow us: Blog | X | LinkedIn | Instagram

Start a free trial today: https://www.crowdstrike.com/trial