WHITE PAPER

SECURING AI SYSTEMS:

A PLAYBOOK FOR SECURITY LEADERS

Summary

Artificial intelligence is transforming the enterprise — embedded in SaaS platforms, powering internal copilots, and enabling productivity across engineering, marketing, legal, customer support, and more. But as organizations accelerate their adoption of generative AI (GenAI), large language models (LLMs), and agentic tools, they're introducing new risks faster than most security programs can track or manage.

AI systems don't behave like traditional applications. They are non-deterministic, access sensitive data, and integrate with platforms that fall outside typical security visibility. Threat actors are already exploiting this gap. Meanwhile, many organizations lack the governance, testing, and policy frameworks to secure AI systems at scale.

This white paper is a guide for security and risk leaders. It examines how AI is reshaping the threat surface, why conventional security thinking falls short, and what practical steps organizations can take to identify, assess, and reduce AI-related exposure. From data governance and regulatory readiness to AI visibility, detection, response, and red teaming, this paper outlines the foundational areas to focus on now — and how to get ahead of the next wave of AI-driven risk.

Table of Contents

AI Adoption Is Accelerating, but Security Isn't

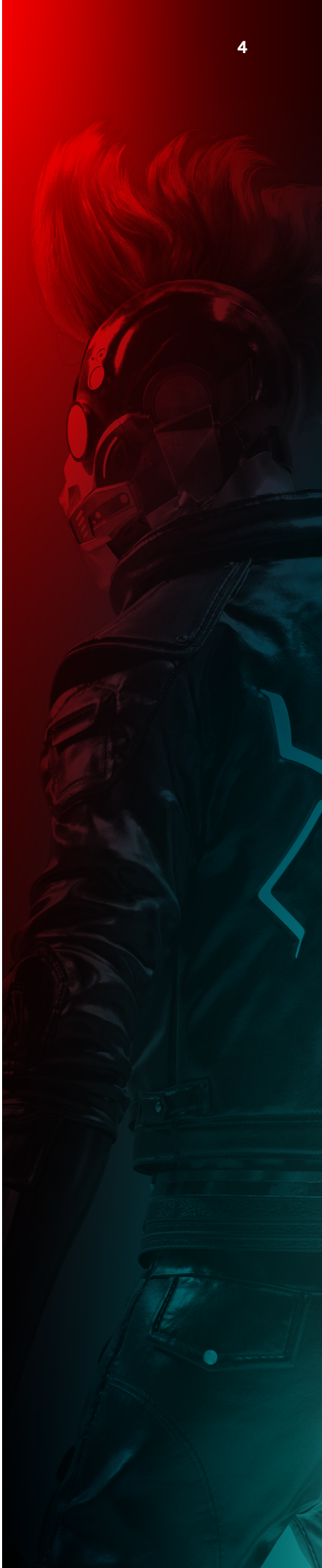
AI adoption is moving faster than any previous enterprise technology wave. From LLM-powered copilots and search assistants to embedded SaaS features and autonomous agents, AI is being rolled out across departments — often without security review or governance.

In most organizations, security teams find out about new AI use cases only after they're in production. Data is flowing through GenAI tools that haven't been assessed, integrations are being enabled without policy enforcement, and agentic behavior is being deployed without monitoring or containment strategies. Security teams have spent years advocating for defense-in-depth strategies — all for copilots and agents to have unfettered and unmanaged access across expansive IT environments.

Moreover, most security personnel lack the experience to understand the unique threats posed by models, AI-powered applications, and agents. Current threat intelligence reveals threat actors using generative AI for productivity: Think deepfakes, AI-generated malware, and AI-generated phishing campaigns. And threat actors have begun to actively exploit AI deployed within victim organizations.[1] We are not far away from talking about prompt injection and data poisoning with the same frequency we discuss ransomware.

This creates a perfect storm: a rapidly growing attack surface that's dynamic, decentralized, and often invisible to the security team. The traditional security model — focused on static infrastructure, known software assets, and clear trust boundaries — doesn't translate cleanly to AI. What makes AI powerful also makes it risky: LLMs and agents operate based on context, respond unpredictably, and interact across systems.

To stay ahead of these risks, organizations must begin with a clear understanding of where AI is being used, what data it touches, and how it behaves in the wild. Without that foundation, there's no way to secure what matters.



1 Anthropic. Nov 13, 2025. https://www.anthropic.com/news/disrupting-AI-espionage
Factory. Nov 20, 2025. https://factory.ai/news/droid-neutralizing-fraud
CrowdStrike 2025 Global Threat Report. https://www.crowdstrike.com/en-us/global-threat-report/

The Challenge of Securing AI Systems

At Fal.Con 2025, CrowdStrike CEO George Kurtz asked the audience during his keynote if anyone had assigned identities to their AI agents. Not a single hand went up. The same security professionals advocating for identity solutions for human and service accounts hadn't yet confronted the reality that agents now need the same treatment. This is because securing AI isn't just a new feature in existing programs — it's a new discipline, a new way of approaching security.

AI systems don't behave like traditional applications. They're probabilistic, context-sensitive, and capable of making independent decisions. This makes conventional approaches to security — focused on known assets, static policies, and rule-based detection — insufficient. Instead, security teams must account for an entirely new set of behaviors, dependencies, and attack vectors.

AI security requires understanding and managing:

- **LLMs** that generate responses based on user prompts and internal context
- **Autonomous agents** that can make decisions, execute tasks, and chain workflows together
- **Third-party AI embedded in SaaS tools** that bypass traditional change management or procurement review
- **Open-source models** deployed internally without formal evaluation or red teaming
- **Externally hosted models** that process sensitive data or connect to enterprise systems

Each of these categories introduces new security questions:

What data is accessible to this model, and is it protected appropriately?

What can the model or agent do, and under whose credentials?

What happens if a user manipulates a prompt or input?

How is this system monitored, logged, and governed over time?

Without clear answers to these questions, organizations face significant blind spots. LLMs can leak sensitive information. Agents can escalate privileges silently. SaaS applications can activate AI features without notifying customers. These are not theoretical concerns — they're happening now, and at scale.

The takeaway is clear: AI systems cannot be secured as an afterthought. They must be treated as dynamic, integrated, and high-risk components of the enterprise architecture — and given the same level of scrutiny as any critical system.

Five Areas to Focus On Now

Securing AI doesn't come down to a single fix or technology. It requires a comprehensive strategy across governance, architecture, visibility, and testing. AI systems behave differently, interact with data differently, and introduce risk in ways that traditional controls can't fully address.

These five areas represent the most urgent domains where security teams must take action — not in isolation, but as part of a cohesive, organization-wide approach to AI risk management:

- Visibility and governance of unsanctioned AI
- Regulatory readiness and compliance
- Data governance: classification, access, and control
- AI detection and response strategy
- Red teaming and adversarial testing of AI systems

1. Visibility and Governance of Unsanctioned AI

AI adoption isn't just coming from the top. It's emerging from the bottom up — led by users, teams, and tools that are moving faster than policy can keep up.

Employees are using AI chatbots and browser extensions powered by LLMs. SaaS platforms are enabling AI features with little to no change control. Developers are plugging open-source models into internal tools and building homegrown AI agents and workloads with insufficient runtime security guardrails. And AI copilots are being rolled out via Microsoft 365, Salesforce, Zoom, and other platforms with little visibility for security teams.

In two recent assessments, CrowdStrike Services discovered hundreds of unaccounted-for agents in production environments. One organization's enterprise AI committee had not approved the use of agentic AI, and the other had an inventory of approved agents that was about 400 short of the number of agents in use. These organizations are different in size, industry, and mission, but they both face the same challenge — implementing AI for the business outpaced governance and security.

WHY SHADOW AI IS HARD TO DETECT

Unlike traditional shadow IT, which typically involves unauthorized devices or SaaS accounts, shadow AI hides in plain sight:

- End users can access external chat interfaces or launch LLM agents with no installation.
- AI features may be bundled into existing tools via license updates.
- Data may be routed through AI services embedded in productivity tools without disclosure.
- AI browser extensions, plug-ins, and third-party copilots operate outside the corporate security stack.
- Development of AI agents and workloads internally may leverage public LLMs without proper data governance and runtime security guardrails.

Security teams cannot rely on traditional asset discovery or traffic inspection alone. Visibility must be rebuilt with AI in mind.

WHAT'S AT RISK

When AI is used without oversight, organizations lose control over:

- **Where their data goes:** Sensitive content may be fed into public LLMs with unknown data retention or training policies.
- **Who can access what data:** Permissions are not persisted when adding contextual data, causing gaps in who can access sensitive data.
- **How agents behave:** Autonomous tools may execute workflows, integrate with APIs, or escalate access with no logging or guardrails.
- **Who's accountable:** Without governance, it's unclear who owns or monitors each instance of AI use.

This not only creates security and compliance risk — it undermines the organization's ability to build a cohesive AI strategy.

HOW TO REGAIN CONTROL

Organizations must operationalize visibility and policy enforcement for AI, just as they did for cloud and SaaS:

- **Establish accountability:** Form a cross-functional team whose mission is to understand AI business needs and help enable them securely.
- **Endpoint and browser discovery:** Use CrowdStrike Falcon® AI Detection and Response (AIDR) to illuminate employee shadow AI adoption and CrowdStrike Falcon® Exposure Management to identify AI components running across your environment, including LLMs, AI agents, IDE extensions, MCP servers, and AI-infused packages.
- **SaaS and cloud discovery:** Use CrowdStrike Falcon® Cloud Security, which includes AI security posture management (AI-SPM); CrowdStrike Falcon® Shield SaaS security posture management; and Falcon AIDR to identify where AI is being used across SaaS platforms, especially in productivity and communication tools.
- **Policy enforcement:** Create clear rules for acceptable use of GenAI tools — both sanctioned and unsanctioned — and publish an approved list of models and interfaces.
- **Agent governance:** Define policies around who can create or deploy AI agents, what those agents can access, what guardrails must be in place, and how agent behavior is logged, reviewed, and deactivated if needed.
- **Awareness and training:** Confirm employees understand the risks of using unauthorized AI tools — including data exposure, policy violations, and integration risk.

Bottom Line:
Shadow AI isn't just a governance problem. It's a visibility and risk problem. To stay ahead of it, security teams need to see, assess, and control AI adoption wherever it occurs — not just where it was planned.

2. Regulatory Readiness and Compliance

AI regulation is no longer theoretical. The EU AI Act has passed and will begin enforcement in staggered phases throughout 2025 and 2026. The U.S., U.K., Singapore, Canada, and others are actively drafting or piloting AI oversight frameworks. Most share common principles: transparency, risk categorization, auditability, and accountability.

For security and risk leaders, this changes the posture around AI adoption. It's no longer just a question of "Can we secure it?" — it's now also "Can we prove we've secured it to regulators, partners, and users?"

To prepare for this shift, organizations must:

- **Inventory AI systems and use cases:** Engage the business in understanding its use cases. Know which models are in use, where they're deployed, and what they're connected to.
- **Classify by risk level:** Map AI systems to emerging categories (e.g., minimal, limited, high-risk) based on use case, data processed, and decision-making authority. This is especially important where regulation, such as the EU AI Act, has different requirements for different risk levels.
- **Support traceability:** Maintain immutable logs for training data, model configuration, prompts, and outputs. This is essential for proving model behavior and root cause analysis in incidents.
- **Codify acceptable use:** Define what use cases are allowed or prohibited, what data can be used, and what vendors must disclose before onboarding AI tools.
- **Engage legal and privacy teams:** Security cannot operate in isolation. Legal, compliance, data governance, and platform teams must co-develop internal guidelines and response protocols for AI systems.

It's also important to note that regulators are increasingly scrutinizing how AI systems make decisions — not just what decisions they make. Security leaders should anticipate pressure to explain model behavior, demonstrate safeguards, and identify where human oversight exists (or is absent). Organizations should also conduct drift and bias monitoring to confirm their AI-enabled systems perform effectively over time.

Bottom Line:
Governance must shift from policy-on-paper to evidence-on-demand.

3. Data Governance: Classification, Access, and Control

AI systems are only as trustworthy as the data that powers them. But in most organizations, data governance hasn't evolved to account for how AI models ingest, process, and expose information. CrowdStrike Services performs hundreds of maturity assessments with customers each year and consistently finds even mature organizations lack basic labeling and data loss prevention controls. The massive amounts of data flowing through AI models compound this issue. Consider a recent use case of a healthcare organization with over 100K employees that launched an enterprise copilot without fully vetting and restricting its access to protected health information. Without clear ownership, classification, and access policies, sensitive data can end up fueling unmonitored models — increasing the risk of leakage, misuse, or regulatory violation.

To secure AI, organizations must treat data governance as a front-line security function. This includes addressing five interdependent domains:

- Data classification and sensitivity
- First-party collection and consent
- Model access and control boundaries
- Training and fine-tuning safeguards
- Operational data segmentation

DATA CLASSIFICATION AND SENSITIVITY

Start with visibility. Know what data your AI systems touch — training data, prompt inputs, knowledge bases, real-time integrations — and classify it accordingly. Label data for:

- Sensitivity — e.g., personally identifiable information (PII), protected health information (PHI), intellectual property (IP)
- Source and ownership
- Regulatory scope
- Acceptable use — e.g., can this be used in production prompts? in model fine-tuning?

Without classification, there can be no meaningful policy enforcement.

FIRST-PARTY COLLECTION AND CONSENT

Understand the provenance of all data flowing into your AI systems. That includes user-submitted inputs, internal datasets, and third-party sources. Legal and compliance teams should validate:

- Whether consent was obtained
- Whether retention is allowed
- What rights are attached to reuse or model training
- Whether data transfer or jurisdictional controls apply

Treat this as part of risk and procurement review — not just privacy governance.

MODEL ACCESS AND CONTROL BOUNDARIES

Many security teams focus on protecting data from AI. But equally important is protecting AI from insecure access. Apply traditional access control principles to AI systems:

- Who can prompt or interact with the model?
- What systems can the model access? (files, APIs, chat history, telemetry)
- Are roles and permissions mapped to identity controls?
- Are logs kept for input/output sessions?

Apply least privilege, Zero Trust principles, and monitoring to model-level access.

TRAINING AND FINE-TUNING SAFEGUARDS

If your organization is training or customizing models, take steps to sanitize and scope what goes in. This includes:

- Removing PII, PHI, and IP
- Excluding sensitive or regulatory-bound data from training sets
- Using filtering pipelines and manual validation during ingestion
- Documenting training data lineage for audit and debugging purposes

Unchecked training data can create downstream exposures that are difficult to detect but easy to exploit.

OPERATIONAL DATA SEGMENTATION

AI environments should never blur the lines between development, staging, and production. Segregate model environments to:

- Prevent test data or insecure scripts from leaking into production
- Protect production prompts, API keys, and output logs
- Ensure red team or testing environments cannot impact live deployments

This is especially critical for environments where LLMs are integrated into chat interfaces, workflows, or decision-making pipelines.

Bottom Line:
Strong data governance isn't just compliance hygiene — it's a prerequisite for AI security. It determines how exposed your data is, how your models behave, and whether you can enforce meaningful controls.

4. AI Detection and Response Strategy

As AI adoption accelerates across the enterprise, security teams face a new detection and response challenge: AI systems operate at runtime, reason in natural language, and interact dynamically with users, data, and downstream systems. Workforce adoption of generative AI tools is often decentralized and opaque, while homegrown models and agents are being deployed directly into production workflows. Together, these trends introduce a rapidly expanding attack surface that traditional detection and response models were never designed to cover: the prompt and agent interaction layer.

Unlike conventional applications, AI systems can be manipulated through prompts, poisoned through data exposure, or coerced into unsafe behavior without triggering classic indicators of compromise. Effective security therefore requires moving beyond static controls and shifting toward continuous detection and response at the AI interaction layer wherever it occurs.

WHY AI DETECTION AND RESPONSE IS DIFFERENT

AI-driven systems introduce several security challenges that complicate detection and response compared to traditional attack surfaces:

- **Language as an attack vector:** Prompts, context windows, and tool calls can be abused to override safeguards, extract sensitive data, and manipulate agent behavior.
- **Non-deterministic behavior:** The same input may yield different outputs, making it difficult to define static "known bad" signatures.
- **Highly distributed execution paths:** AI interactions span browsers, SaaS platforms, APIs, gateways, agents, and cloud workloads.
- **Invisible misuse:** Employees and agents may unintentionally expose data or trigger unsafe actions during otherwise legitimate workflows.

Without dedicated visibility into AI interactions at runtime, organizations are left blind to many of the most likely AI-specific threats.

BUILDING AN EFFECTIVE AI DETECTION AND RESPONSE STRATEGY

Security leaders should approach AI detection and response as a first-class discipline, aligned to the realities of how AI is used and attacked in practice. Key strategic elements include:

- **Unified runtime visibility:** Organizations must observe AI activity where it actually occurs: across workforce tools and internally developed AI systems. This includes capturing prompts, responses, agent actions, model context, and integrations to establish a baseline of normal behavior and enable investigation when anomalies arise.
- **Behavior-based threat detection:** Detection should focus on identifying malicious or risky patterns such as prompt injection, indirect manipulation via retrieved content, unsafe tool invocation, excessive agency, and anomalous access to sensitive data. These techniques often evade traditional security controls because they operate at the semantic rather than technical layer.
- **Real-time data protection:** Sensitive information — including credentials, regulated data, proprietary code, and intellectual property — must be protected before it reaches an external model or is exposed in AI-generated output. Runtime inspection and transformation response controls are critical for preventing irreversible data leakage.
- **Policy-driven response and containment:** AI security controls must operate at machine speed. When unsafe behavior is detected, organizations should be able to block interactions, redact or encrypt sensitive data, restrict access based on context, and halt agent execution without disrupting legitimate use.
- **SOC integration and investigability:** AI detections should feed directly into existing security operations workflows and SIEM environments. Rich telemetry — including full interaction context — is essential for triage, root cause analysis, and continuous improvement of AI security controls.

FROM REACTIVE CONTROLS TO CONTINUOUS DEFENSE

AI systems cannot be secured through policy alone. Detection and response capabilities specific to AI must evolve alongside how models are used, trained, and integrated. By instrumenting the AI runtime layer, establishing robust detections, and enforcing automated response, organizations can enable safe AI adoption and deployment while reducing the risk of misuse, compromise, and data exposure.

Bottom Line:
AI expands the attack surface in ways that bypass traditional detection and response controls. A dedicated AI detection and response strategy gives security teams the visibility, context, and control they need to defend AI systems — without slowing innovation or adoption.

5. Red Teaming and Adversarial Testing of AI Systems

Traditional penetration testing is not equipped to evaluate how AI systems behave under adversarial pressure. While infrastructure, identity, and network controls may be secure, the model itself introduces an entirely different layer of risk.

AI systems are non-deterministic and contextual. Their behavior can change based on prompt history, user identity, or external plug-ins. This makes them highly exploitable in ways that evade traditional security review.

During testing of deployed AI-integrated systems, CrowdStrike Services has demonstrated attacks such as RAG poisoning for data extraction and code execution, jailbreaking of an agentic system that led to code execution and credential theft, and causing models to provide false legal and financial advice. Traditional vulnerability scanning struggles to discover these types of attacks, and so will red teamers without deep AI understanding.

To properly evaluate risk, organizations must simulate how adversaries interact with AI systems in the real world. That means moving beyond infrastructure scanning to model-aware, behavior-driven testing.

KEY ADVERSARIAL TESTING TECHNIQUES

Effective red teaming for AI systems should include:

- **Prompt injection attacks:** Manipulate the model's behavior using crafted inputs to bypass instructions or unleash unintended functionality.
- **Data leakage and model inversion:** Attempt to extract sensitive information from training data, previous sessions, or integrated knowledge bases through carefully engineered queries.
- **Output manipulation:** Coerce the model into generating harmful or misleading outputs — including instructions, impersonation, or social engineering content.
- **Function call and plug-in abuse:** If the model is connected to external tools, APIs, or internal plug-ins, testing should include abuse of those integrations to escalate privileges or exfiltrate data.
- **Agentic behavior validation:** For multi-step agents, red teams must test for unintended task chaining, recursive loops, and unsafe decision-making logic.

FRAMEWORKS AND METHODOLOGIES

Organizations should align testing efforts with established frameworks, such as:

- **MITRE ATLAS™** for adversarial tactics and threat modeling against AI and machine learning systems
- **OWASP Top 10 Risks and Mitigations for LLMs and Gen AI Apps**
- **OWASP Top 10 Risks and Mitigations for Agentic Applications**
- **AI risk taxonomies** to prioritize model- and data-specific attack surfaces, such as CrowdStrike's comprehensive Prompt Injection Taxonomy

Red team testing should also be coordinated with detection engineers and SOC analysts to evaluate how well systems respond to attack — including logging, alerting, and containment workflows.

Bottom Line:
AI systems are attractive targets. They contain valuable data, influence decision-making, and increasingly serve as the interface between users and critical operations. But they also behave unpredictably — and attackers know it. If you're not actively testing how your AI systems can be abused, you're relying on assumptions — not evidence.

Securing AI Requires a Threat-Informed Approach

For years, organizations have approached governance through policy and process — acceptable use guidelines, procurement checklists, and compliance audits. While important, these tools alone are not sufficient in the age of AI.

Why? Because threat actors don't care about your policies. They exploit behavior, misconfigurations, and blind spots — especially in technologies that are new, fast-moving, and inconsistently governed.

AI systems fit that description perfectly.

To secure AI effectively, organizations must shift from static risk models to **threat-informed security** — a mindset that asks: How would an attacker break this system? And how can we stop them before they do?

Principles of a Threat-Informed Approach

1. **Think like an adversary**

- Map how attackers might interact with your AI systems
- Identify the trust boundaries between inputs, models, outputs, and downstream systems
- Understand how LLMs, agents, and plug-ins might be manipulated, bypassed, and escalated

2. **Red team the model, not just the perimeter**

- Simulate adversarial interactions directly against the LLM or agent
- Test for prompt injection, model override, function abuse, and information leakage
- Include both authenticated and unauthenticated attack paths

3. **Instrument telemetry and response**

- Monitor model activity: user prompt logs, additional context, LLM responses, agent tool calls, frequency patterns, and escalation paths
- Establish baselines for "normal" model behavior — then alert on anomalies
- Integrate AI-specific detections into the SOC and the broader incident response plan

4. **Validate controls through simulation**

- Use red teaming to test both model response and detection logic
- Track whether output filtering, API gating, and policy controls hold up under realistic attack conditions
- Close the loop between adversary simulation and mitigation

Make It Continuous

A threat-informed approach is not a one-time assessment. AI models evolve. Usage patterns shift. New features emerge.

Security must evolve alongside them — with real-time visibility, ongoing testing, and continuous collaboration between security, engineering, and platform teams.

Securing AI isn't just about building fences. It's about understanding how your models behave under pressure — and designing your controls, monitoring, and response plans with that behavior in mind.

Get Ahead of the Risk

AI is no longer an emerging trend — it's an operational reality. Models, agents, and AI-enhanced platforms are being deployed across the enterprise, often without visibility, governance, or safeguards in place.

This wave of adoption introduces new and dynamic risks:

- Sensitive data moving through opaque models
- Misconfigurations and integrations outside security review
- Agents capable of taking action without auditability
- Threat actors actively exploiting these weaknesses

Security leaders can't wait for AI risk to be handed to them in a clean package. They must proactively identify where AI is being used, how it behaves, and what needs to change to make it safe.

This means:

- Discovering and inventorying all AI usage across the environment
- Enforcing visibility and policy around shadow AI and unauthorized usage
- Classifying the data AI systems access, generate, and store
- Assigning ownership and establishing cross-functional governance
- Testing model behavior through red teaming and adversary simulation
- Building detection and response workflows tailored to AI-specific threats

Above all, it means shifting from reactive posture to proactive defense — using a threat-informed, evidence-driven approach to secure AI systems as they become core to business operations.

Organizations that succeed won't just secure AI. They'll accelerate its adoption and deployment.

How CrowdStrike Can Help

CrowdStrike provides expert-led consulting to help organizations identify, assess, and reduce the risks introduced by artificial intelligence. Whether you're just starting your AI journey or already deploying LLMs, copilots, and agentic tools at scale, CrowdStrike experts provide the strategy, testing, and visibility to secure your AI footprint with confidence.

Here's what's included in the portfolio.

CrowdStrike Falcon Platform

FALCON AI DETECTION AND RESPONSE (AIDR)

Protect your AI investments with comprehensive visibility, detection, and automated response capabilities that secure AI systems from emerging threats like prompt injection and data leakage. Falcon AIDR helps security teams illuminate shadow AI and runtime behavior, detect sophisticated AI-specific attacks, and enforce granular access controls and respond automatically — all managed through a single console within the same trusted platform that protects your endpoints, cloud, and identities.

Customers receive:

- Visibility into AI use and behavior across workforce and applications with a unified AI activity graph
- Industry-leading prompt injection detection with up to 99% efficacy and sub-30ms latency[2]
- Automated response actions to block attacks, mask sensitive data, and enforce AI governance policies
- Seamless deployment through lightweight browser extensions, software development kit (SDK) integration, AI/API gateway integrations, MCP proxies, and cloud integrations.

CrowdStrike also provides additional platform-native tools to continuously secure AI environments, including:

- **Falcon Shield:** Discover AI agents across SaaS platforms like Microsoft 365, Salesforce, and OpenAI. Map their system access, detect risky behavior, and contain threats — before AI automation turns into exploitation.
- **Falcon Exposure Management:** Identify and secure AI components running across endpoint and cloud environments, including LLMs, AI agents, IDE extensions, MCP servers, and AI-infused packages.
- **Falcon Cloud Security — AI-SPM and Image Assessor:** Discover, assess, and secure AI usage in cloud and containerized environments.
- **CrowdStrike Falcon® Next-Gen Identity Security:** Secure AI agents and the identities behind them, often created without security oversight. Gain deep visibility into usage across platforms like Microsoft 365, OpenAI, and Snowflake. Uncover unsanctioned access and risky behavior, and stop adversaries from hijacking automation to escalate privileges, modify code, or exfiltrate data.

2 Performance metrics are based on results from CrowdStrike internal benchmark testing.

CrowdStrike Professional Services

Together, these services help organizations:

- Identify where AI is in use
- Assess and reduce exposure
- Validate model behavior under adversarial pressure
- Build governance and response strategies grounded in real-world threats

AI SYSTEMS SECURITY ASSESSMENT

Gain visibility into how AI is being used across your organization — including unsanctioned AI, third-party copilots, and unsafe integrations. This service identifies high-risk model usage, misconfigurations, and access control gaps across SaaS, cloud, and internal environments.

Customers receive:

- Discovery powered by Falcon Shield and Falcon Cloud Security's AI-SPM
- Technical analysis of AI pipelines, model exposure, and integrations
- Strategic workshops and a prioritized roadmap to reduce AI risk
- Model evaluation using LLM Arena for prompt injection and configuration testing

AI RED TEAM SERVICES

Adversarial testing is essential for validating the security of AI systems. This service simulates real-world attacker tactics against models, copilots, and agents to uncover issues before adversaries do.

Testing includes:

- Prompt injection and output manipulation
- Model evasion and inversion attacks
- Abuse of plug-ins, function calling, and agent chaining
- Threat modeling aligned to MITRE ATLAS and OWASP Top 10 Risks and Mitigations for LLMs

AI FOR SECOPS READINESS

Many organizations are being asked to bring AI into their security operations — but they don't have a clear plan. This service helps define the right use cases, assess readiness, and design a secure implementation strategy.

Customers receive:

- SOC workflow review and AI adoption strategy
- Use case design and prioritization
- Architecture planning and "build vs. buy" guidance
- Reference designs for secure GenAI adoption across the detection and investigation chain to increase the speed, scale, and accuracy of the response

Start securing your AI footprint today.

Learn more at https://www.crowdstrike.com/en-us/solutions/secure-your-ai/
or reach out to services@crowdstrike.com

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: www.crowdstrike.com

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

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