Solution overview Cisco public

AI Defense

Speed, Security, and Safety: winning the race to embrace AI

As a security leader, you're facing an AI revolution at light speed. AI is now table stakes for competitive advantage, forcing you into a delicate balancing act: move fast or fall behind, but not at the expense of security and safety.

AI teams are rushing to deploy new AI-enabled applications. Consequently, your IT infrastructure is evolving into a multi-cloud, multi-model, multi-stakeholder AI ecosystem–one you need to secure.

Securing AI requires a new approach

AI disrupts traditional security models. Unlike static software, AI is non-deterministic, often exhibiting unexpected behaviors that can expose sensitive data or generate harmful advice. These evolving behaviors introduce AI-specific threats, including data poisoning, model manipulation, and adversarial attack–risks existing security controls fail to detect. With AI, security (e.g., protecting against prompt injections), privacy (e.g., preventing Protected Health Information [PHI] leaks), and safety (e.g., preventing AI from advising self-harm) are equally critical. In this evolving landscape, three crucial challenges emerge for security teams:

1. **Visibility gaps:** You need a clear view of every AI application, model, and dataset operating across multiple clouds. As with other areas of multi-cloud security, AI demands specialized third-party solutions to monitor and manage these shared environments.
2. **AI model and application vulnerabilities:** To maintain compliance and manage risk, you must continually assess your AI models and applications around how susceptible they are to various safety and security risks.
3. **Emerging adversarial AI threats:** AI systems are vulnerable to novel attack vectors targeting runtime applications (e.g., prompt injections and jailbreaking). Your existing cybersecurity measures are oblivious to these threats.

By proactively taking on these challenges, Cisco® AI Defense empowers you and your security team to ensure safe and secure AI adoption. Building on your existing network visibility and enforcement controls, AI Defense embeds pioneering, industry-recognized AI and cybersecurity technologies that deliver the oversight, protection, and security you need to develop, deploy, and use AI–without sacrificing security, safety, or speed.

Cisco AI Defense Benefits

- Enable secure, private, and safe AI-enabled app development with real-time validation of models and assets through algorithmic AI red teaming.
- Defense against sensitive data loss and advanced adversarial threats with leading threat intelligence and AI guardrails.
- Accelerate value with a single management plane for all AI and smooth integration with the Cisco Security Cloud.

Primary AI Defense use cases

- **Discover** – Maintain continuous visibility into AI-related cloud traffic (ingress, egress, east-west) while automatically discovering AI assets like models and agents.
- **Detect** – Proactively test and validate all AI models—open-source and proprietary—to detect vulnerabilities early and safeguard your AI environment continuously.
- **Protect** – Protect your AI infrastructure, including agent-based AI workflows and retrieval augmented generation (RAG) applications from malicious prompts, data leakage, and adversarial attack.

Core components and capabilities

Securing AI requires a holistic approach that protects your enterprise AI-enabled applications. With AI becoming pervasive across your enterprise, legacy solutions leave critical gaps that threat actors can exploit. To mitigate AI risks, your security strategy must be comprehensive, scalable, and built for AI's unique challenges.

As shown in Figure 1, a holistic AI security and safety program combats AI risk across the development of AI applications.

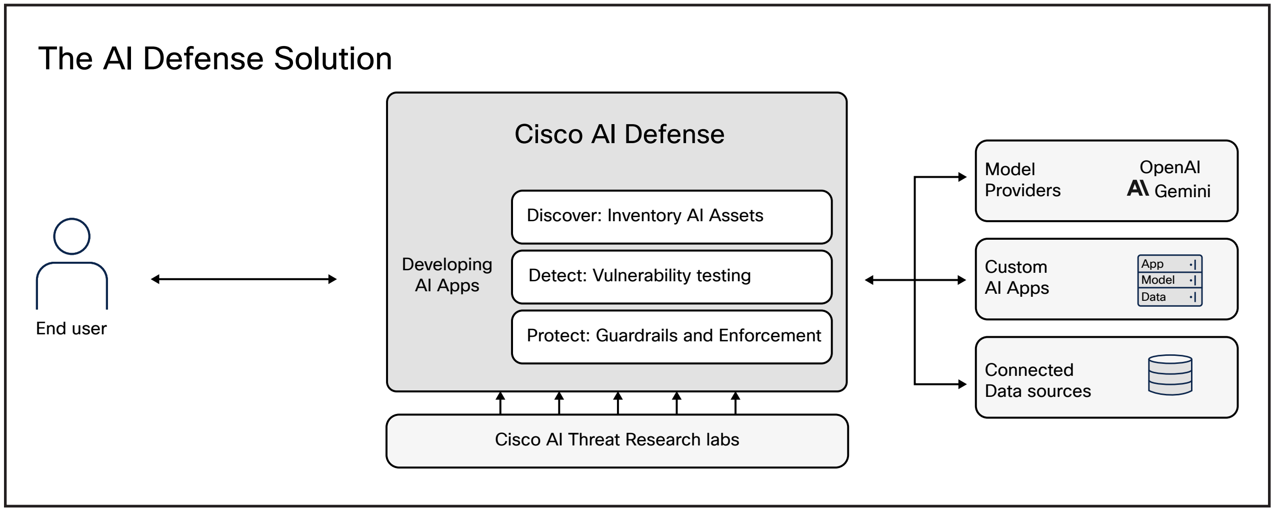


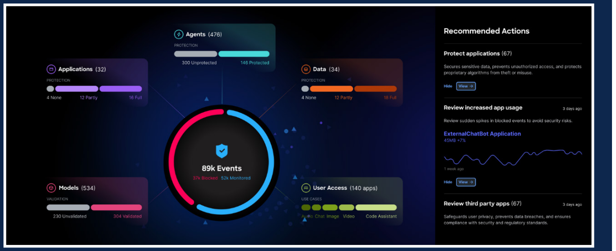
Figure 1. AI Defense

An integrated AI security, privacy, and safety solution ensuring the security, privacy, and safety of your AI footprint requires addressing visibility gaps, AI model vulnerabilities, and the novel AI threat landscape. To do this, AI Defense provides three core components that align with the primary use cases: AI Cloud Visibility (Discover), AI Model and Application Validation (Detect), and AI Runtime Protection (Protect).

Safeguarding enterprise AI with Cisco AI Defense

The AI Defense management console drives a risk dashboard, delivering real-time insights:

- Enterprise AI application usage, context, and risks.
- Discovery, validation, and assessment of the AI models, agents, and related assets (e.g., repositories).
- Runtime defense with security, safety, and privacy guardrails.



Recommended Actions

Protect applications (67)

Secures sensitive data, prevents unauthorized access, and protects proprietary algorithms from theft or misuse.

Review increased app usage 3 days ago

Review sudden spikes in blocked events to avoid security risks.

Review third party apps (67) 3 days ago

Safeguards user privacy, prevents data breaches, and ensures compliance with security and regulatory standards.

AI Cloud visibility

- Prevent security gaps by gaining complete visibility of the enterprise AI attack surface (ingress, egress, and east-west traffic).
- Uncover AI assets across your cloud.
- Make informed decisions about security policies by mapping connections, activity, and identities across data, models, and agents.

AI Model and application validation

- Detect in seconds what takes a manual red team weeks. Ensure application safety and security with algorithmic red teaming to test AI applications and models against 200+ attack techniques and threat categories.
- Generate deep insights from comprehensive vulnerability reports with actionable guardrail recommendations.
- Assess AI models and applications with the latest techniques and threats discovered by Cisco's security teams.

AI Runtime protection

- Protect production AI applications from adversarial attacks, data leaks, compromised resources, and harmful responses.
- Implement prescriptive runtime guardrails to prevent data leaks, threats, and safety violations through prompts and responses.
- Achieve high-accuracy detections using Cisco's proprietary AI models to identify and assess a broad range of threat categories, such as prompt injections, data privacy, security, and safety attacks.

Applying discover, detect, and protect to safeguard AI applications

The AI Defense components seamlessly integrate to discover, detect, and protect AI applications across their entire lifecycle.

AI Cloud visibility	AI Model and application validation	AI Runtime protection
Discover	Detect	Protect
Automatically discover AI assets: ▪ AI Agents and RAG ▪ LLMs ▪ Repositories	**Detect real time AI model and app Risks:** ▪ Adversarial attack susceptibility ▪ Safety risks ▪ Security vulnerabilities	**Implement runtime guardrails to defend against:** ▪ Denial of Service (DoS) ▪ Harmful advice ▪ Prompt injections ▪ Sensitive data leakage
Continually discover new AI assets across your cloud infrastructure.	Configure the validation service to run AI algorithmic red teaming against discovered AI models and apps.	The output of the validation service generates policies (guardrails) that you can implement to drive the runtime protection of Ai applications.
Track and manage risk: AI Cloud visibility lets you understand ownership and context of AI assets.	**Manage compliance and vulnerabilities:** AI model and Application validation continually evaluates vulnerabilities, especially over time as models/apps change.	**Ensure safety and security:** As developers fine-tune AI-enabled apps, they often unintentionally break guardrails. AI defense detects and prevents these failures.
Achieve continuous adversarial threat defense Cisco's AI threat research team supports discover, detect, and protect activities by continually updating AI defense with the latest adversarial attack data aligning with MITRE adversarial threat landscape for Aritificial-intelligence System (ATLAS), OWASP, NIST, and other frameworks.		

Cisco Security for AI

Cisco is building on decades of leadership in networking and cybersecurity to pave the way for rapid AI innovation and resilient AI security. We cover areas across:

- **AI apps built by the enterprise:** Cisco AI Defense protects against the safety and security risks introduced by the development and deployment of AI applications.
- **Third-party GenAI apps or Shadow AI:** Cisco safeguards organizations from the security risks of third-party AI applications with Cisco Secure Access, protecting against threats and sensitive data loss while restricting employee access to unsanctioned tools.
- **AI supply chain and risk management:** Cisco protects against malicious AI model files (e.g., scanning for malicious code, software license compliance, and geopolitical origin risks) entering the enterprise through network-based enforcement with Cisco Secure Access, Cisco Secure Endpoint, and Cisco Email Threat Defense.

Safely innovating at the speed of AI

AI is advancing fast, and your organization wants to keep pace—without compromising security, privacy, or safety. Cisco AI Defense empowers you and your team to confidently deliver AI-driven solutions, blending protection and coverage with the agility your teams need to stay ahead. Here's how:

- **Resilient coverage with security, privacy, and safety:** With Cisco AI Defense, you gain end-to-end coverage throughout the AI application lifecycle—combining automated, algorithmic vulnerability testing with robust runtime guardrails for production applications. You can seamlessly detect, validate, and enforce AI security by leveraging network-layer visibility across the Cisco Security Cloud.
- **Proactive threat landscape navigation:** Stay one step ahead of emerging AI threats with continuous updates from Cisco's AI threat intelligence research groups, global leaders in AI threat tactics, techniques, and procedures (TTPs). AI Defense's real-time intelligence ensures you can anticipate new attack vectors—shoring up your defenses before adversaries can take advantage.
- **Keep your AI Initiatives current:** AI Defense aligns with evolving standards like the National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI-RMF), MITRE Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS), and the Open Worldwide Application Security Project (OWASP) Top 10 for Large Language Models (LLM). Better still, the AI Defense team actively contributes to developing these frameworks—ensuring you're always ready for the next wave of regulatory and industry requirements.

By integrating Cisco AI Defense's discover, detect, and protect use cases into your AI strategy, you equip your organization to push the boundaries of innovation—safely, swiftly, and with complete confidence in your security risk posture.

Next steps

Learn more at https://www.cisco.com/go/ai-defense.