Securing Artificial Intelligence Applications

PointGuard AI is the first application vendor to tackle the challenges of protecting AI applications and GenAI LLM systems. By extending its robust platform, PointGuard AI can integrate management of AI security risks into established application security processes.

Businesses across sectors are seeing the potential of GenAI, experimenting with use cases, and in many cases, already rolling out initial public-facing applications. But in the rush to deployment, it's easy to forget that GenAI and LLM applications introduce significant new risks and new players such as data scientists, who are not well versed in security. At the same time, many security professionals are only vaguely aware of fast-moving AI projects.

Security Challenges

AI applications introduce significant new security risks with a greatly expanded attack surface. These risks include:
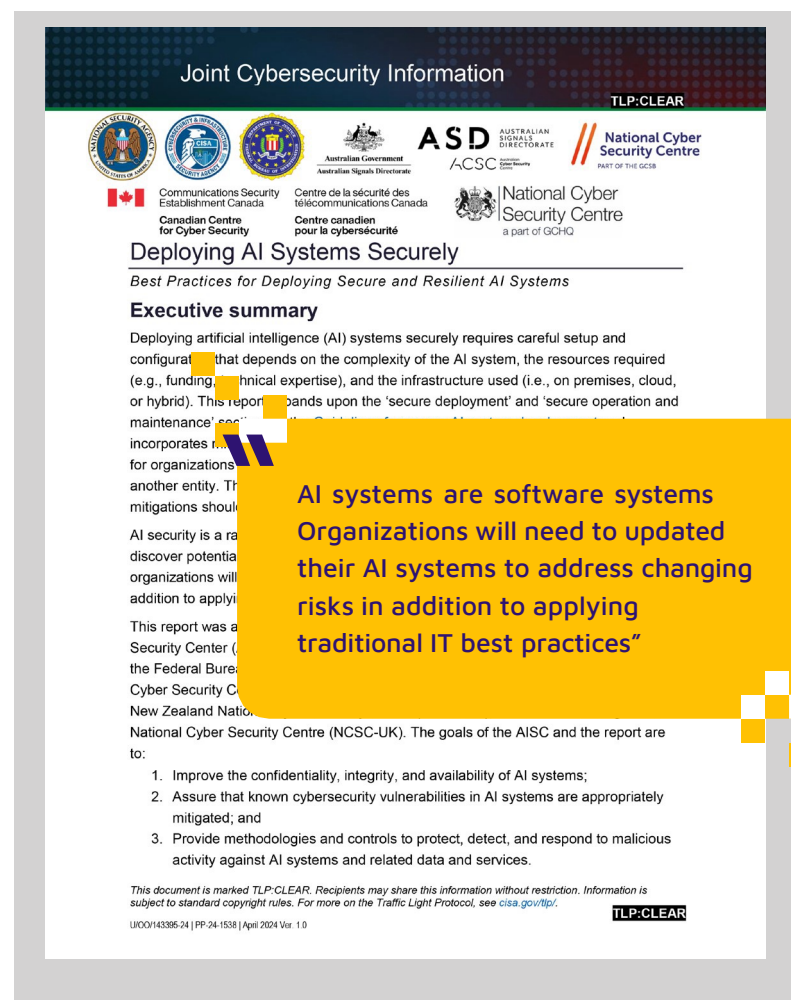
Lack of visibility into AI LLM stacks

- Widespread unsanctioned use without oversight
- Decisions driven largely by Data Science teams
- AI is integrated with core applications but treated as isolated

Evolving risk of AI sprawl

- Rapid adoption of AI with limited focus on security
- LLM vulnerabilities pose risk to connected applications
- Possible leaks of sensitive data

Limited compliance

- Licensing, ownership, IP and regulatory issues
- Lack of lineage or provenance of models
- No governance on training data (GDPR/CCPA, etc.)

Joint Cybersecurity Information

TLP:CLEAR

Deploying AI Systems Securely

Best Practices for Deploying Secure and Resilient AI Systems

Executive summary

Deploying artificial intelligence (AI) systems securely requires careful setup and configuration that depends on the complexity of the AI system, the resources required (e.g., funding, technical expertise), and the infrastructure used (i.e., on premises, cloud, or hybrid). This report expands upon the 'secure deployment' and 'secure operation and maintenance' sections of the Guidelines for secure AI system development and incorporates mitigation considerations from Engaging with Artificial Intelligence (AI) for organizations using AI systems designed and developed by another entity. The mitigations should be...

AI security is a rapidly evolving area of research. As organizations discover potential vulnerabilities and subsequent mitigations, organizations will need to...

> "AI systems are software systems Organizations will need to updated their AI systems to address changing risks in addition to applying traditional IT best practices"

This report was authored by the Australian Signals Directorate's Australian Cyber Security Center (ASD ACSC), the U.S. National Security Agency (NSA), the Federal Bureau of Investigation (FBI), and the following partners: the Cyber Security Centre (CCCS), New Zealand National Cyber Security Centre (NCSC-NZ), the United Kingdom National Cyber Security Centre (NCSC-UK). The goals of the AISC and the report are to:

1. Improve the confidentiality, integrity, and availability of AI systems;
2. Assure that known cybersecurity vulnerabilities in AI systems are appropriately mitigated; and
3. Provide methodologies and controls to protect, detect, and respond to malicious activity against AI systems and related data and services.

This document is marked TLP:CLEAR. Recipients may share this information without restriction. Information is subject to standard copyright rules. For more on the Traffic Light Protocol, see cisa.gov/tlp/.

U/OO/143395-24 | PP-24-1538 | April 2024 Ver. 1.0

TLP:CLEAR

Guidance on Deploying AI Systems Security from the NSA, FBI, and global cyber agencies

Emerging Control Frameworks

Standards organizations are concerned about AI application risks and developing new frameworks to categorize, track, and remediate against threats. These include:

- OWASP Top 10 for LLM Applications

- MITRE ATLAS™ (Adversarial Threat Landscape for AI Systems

OWASP Top 10 for LLM Applications

LLM01	**LLM02**	**LLM03**	**LLM04**	**LLM05**
Prompt Injection	Insecure Output Handling	Training Data Poisoning	Model Denial of Service	Supply Chain Vulnerabilities
Manipulates LLMs through input. Direct injections overwrite system prompts. Indirect manipulate inputs from external sources.	Outputs accepted without scrutiny. Can lead to XSS, CSRF, SSFR, or remote code execution.	Tampered data can introduce vulnerabilities and biases that compromise security.	Resource-heavy operations degrades service, increasing resource-intensive nature of LLMs and user inputs.	Vulnerable components, services, datasets, pre-trained models, and plugins can compromise LLM application lifecycle.
LLM06	**LLM07**	**LLM08**	**LLM09**	**LLM10**
Sensitive Info. Disclosure	Insecure Plugin Design	Excessive Agency	Overreliance	Model Theft
Confidential data inadvertently revealed through responses. Can cause privacy violations and security breaches.	Insecure inputs and insufficient access control can lead to exploits like remote code execution.	Excessive functionality, permissions, or autonomy may lead to unintentional consequences.	Systems or people overly dependent on LLMs may face misinformation, legal issues or vulnerabilities from incorrect LLM-generated content.	Unaithorized access copying, or exfiltration of proprietary LLM models can cause access to sensitive information.

**Threats categorized by OWASP for LLM Applications.
Yellow boxes are addressed by PointGuard AI.**

What's Required

By extending its application security platform PointGuard AI uniquely protects both core applications and new AI applications that are connected. This includes:

- AI application visibility

- LLM supply chain security

- AI stack misconfigurations

- Infrastructure vulnerabilities
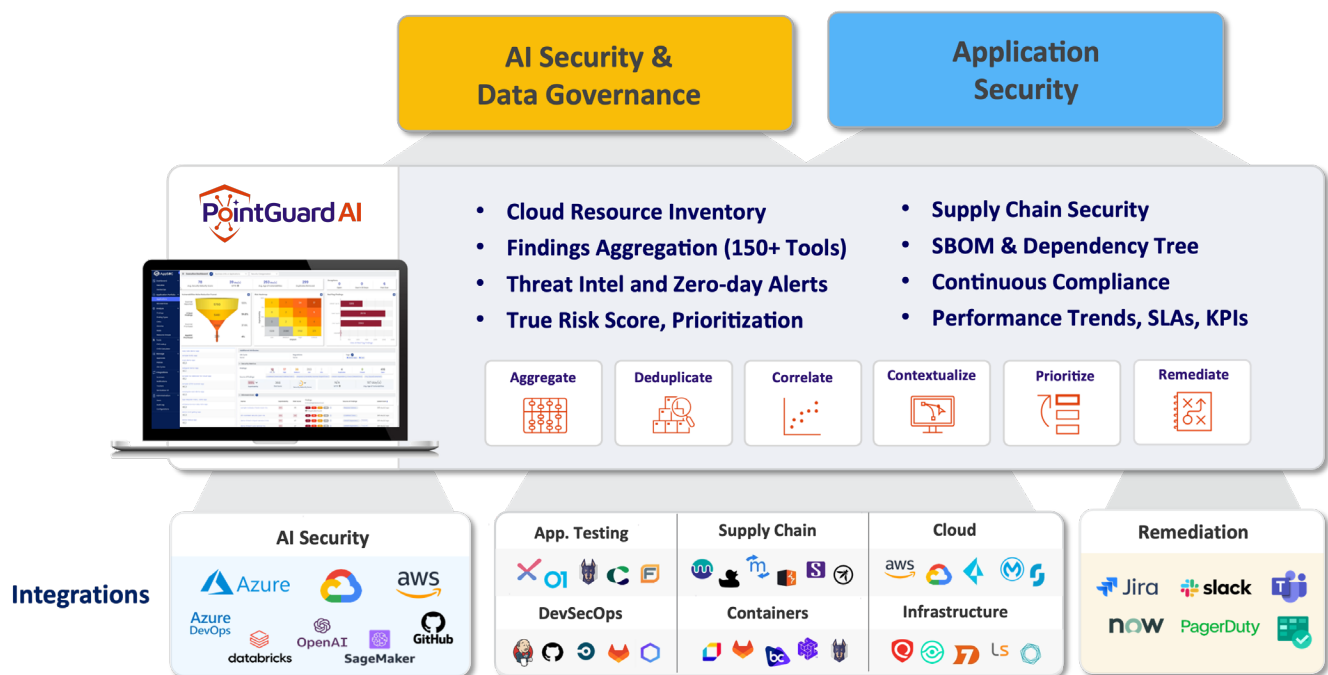
- Data loss and poisoning prevention

PointGuard AI Application Security & Governance

Extending the PointGuard Platform

PointGuard has extended its comprehensive ASPM platform with four new modules addressing AI application risks, including:

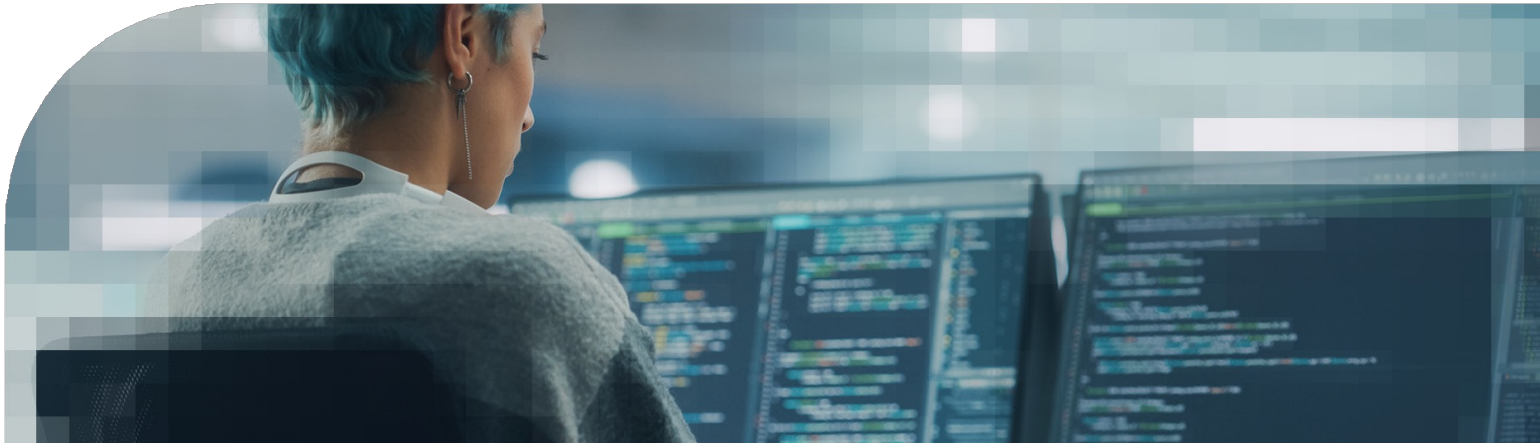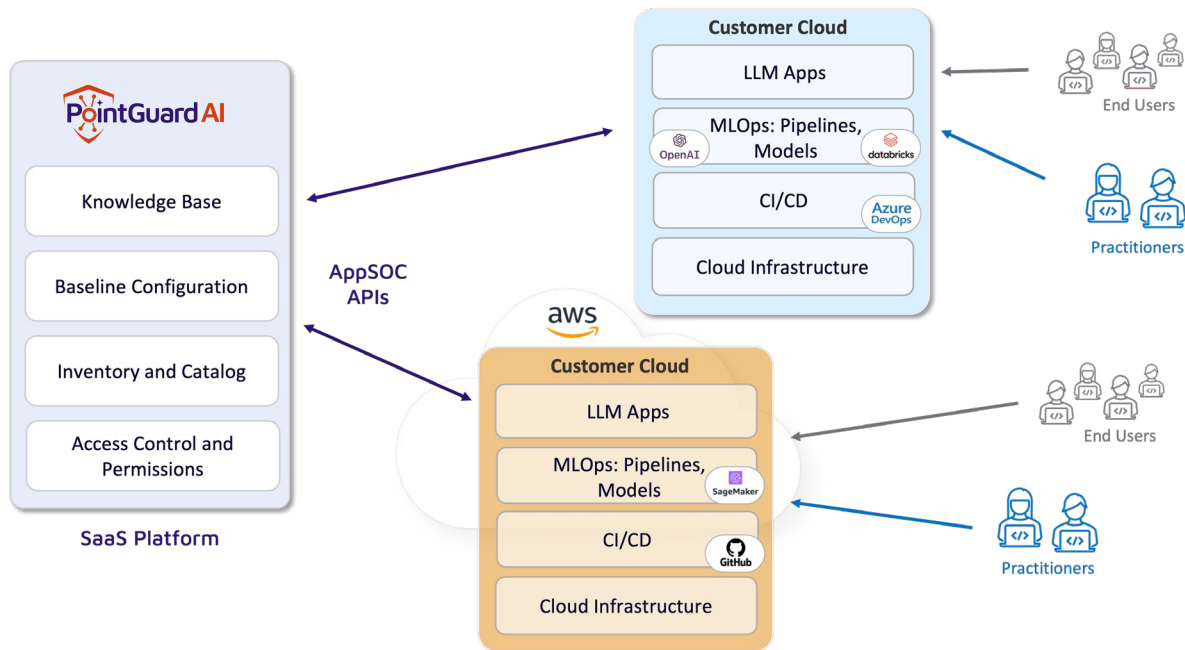| AI Governance	AI Security Posture Management	AI Application Security	Data Protection & Compliance
• Shadow AI discovery • AI use-case repository • Ownership & policy mapping • AI impact assessments	• AI stack visibility (models, integrations, consumers) • Resource inventory • Configuration hardening • KB & risk rating	• Operational risk scanning (security & licensing) • Model scanning • Prompt injection detection • Runtime policy enforcement (bias, acceptable use)	• Data classification & lineage • DLP (PCI/PHI/PII) • Data access anomalies • Identity & application mapping

Frictionless Deployment

The PointGuard AI platform is deployed in the cloud and provides custom API integration with major cloud providers including Microsoft Azure, AWS, and Google Cloud. The AI Application modules integrate with MLOps pipeline tools (including Databricks, OpenAI and SageMaker) as well as CI/CD tools (including Azure DevOps and GitHub). This provides broad security coverage while minimizing the impact on user or customers through cumbersome agents

Enabling AI Initiatives

Blocking important AI projects will not work and will put your company at a competitive disadvantage.

With PointGuard you can deploy AI applications and LLM systems with confidence while ensuring visibility, security, governance, and continuous compliance for all your applications and infrastructure.

PointGuard AI is a leader in Application Security Posture Management (ASPM) and Unified Vulnerability Management (UVM). Our mission is to break through security silos, consolidate data across hundreds of tools, prioritize findings based on real business risk, and reduce the friction between DevSecOps teams to make security more precise and cost-effective. **PointGuard's** global team is headquartered in California in the heart of Silicon Valley.

Learn more at PointGuardAI.com

Follow us on