



# Securing Artificial Intelligence Applications

AppSOC is the first application vendor to tackle the challenges of protecting AI applications and GenAI LLM systems. By extending its robust platform, AppSOC 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

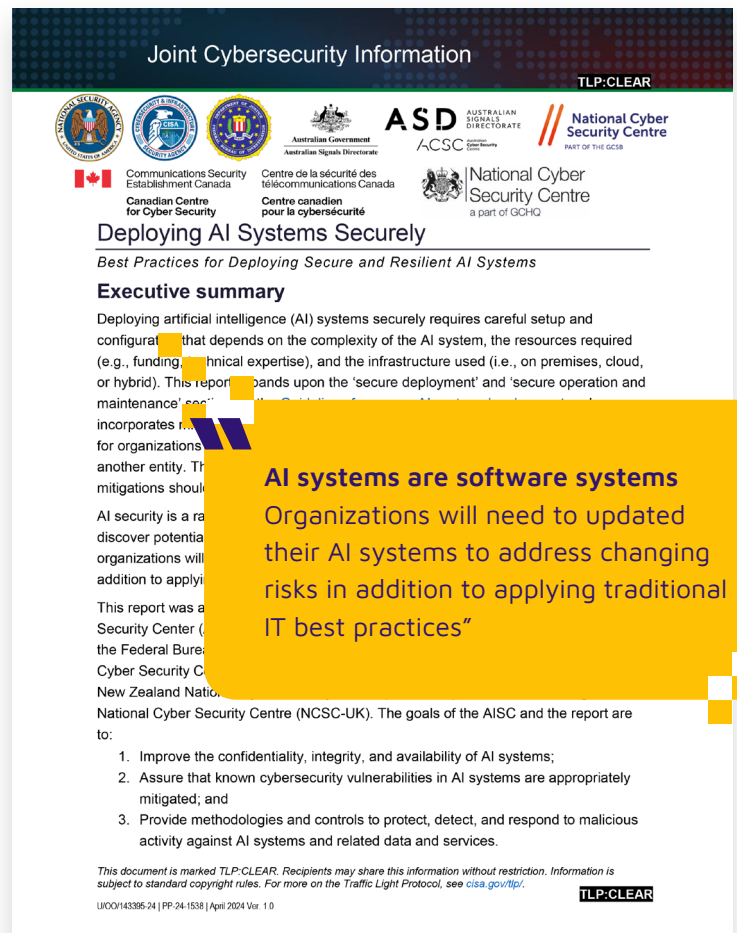
- Widespread unsanctioned use without oversight
- Decisions driven largely by Data Science teams
- AI is integrated with core applications but treated as isolated

### 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 builds upon the 'secure deployment' and 'secure operation and maintenance' components of the AI Security Framework. The report incorporates recommendations for organizations that are not another entity. The mitigations should be applied to AI security is a rapidly evolving field. Organizations will need to discover potential vulnerabilities in addition to applying traditional IT best practices."

This report was a joint effort of the Security Center (NSA/CSS), the Federal Bureau of Investigation (FBI), the Cyber Security Centre (CSC), the New Zealand National Cyber Security Centre (NCSC-NZ), and the 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/](https://cisa.gov/tlp/).

U00143395-24 | 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

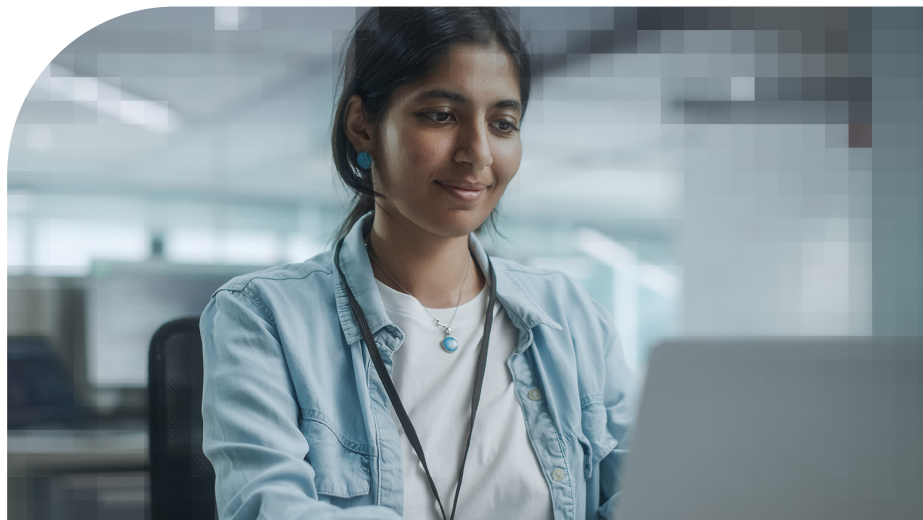
<b>LLM01</b> <b>Prompt Injection</b> Manipulates LLMs through input. Direct injections overwrite system prompts. Indirect manipulate inputs from external sources.	<b>LLM02</b> <b>Insecure Output Handling</b> Outputs accepted without scrutiny. Can lead to XSS, CSRF, SSFR, or remote code execution.	<b>LLM03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

## What's Required

By extending its application security platform AppSOC 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

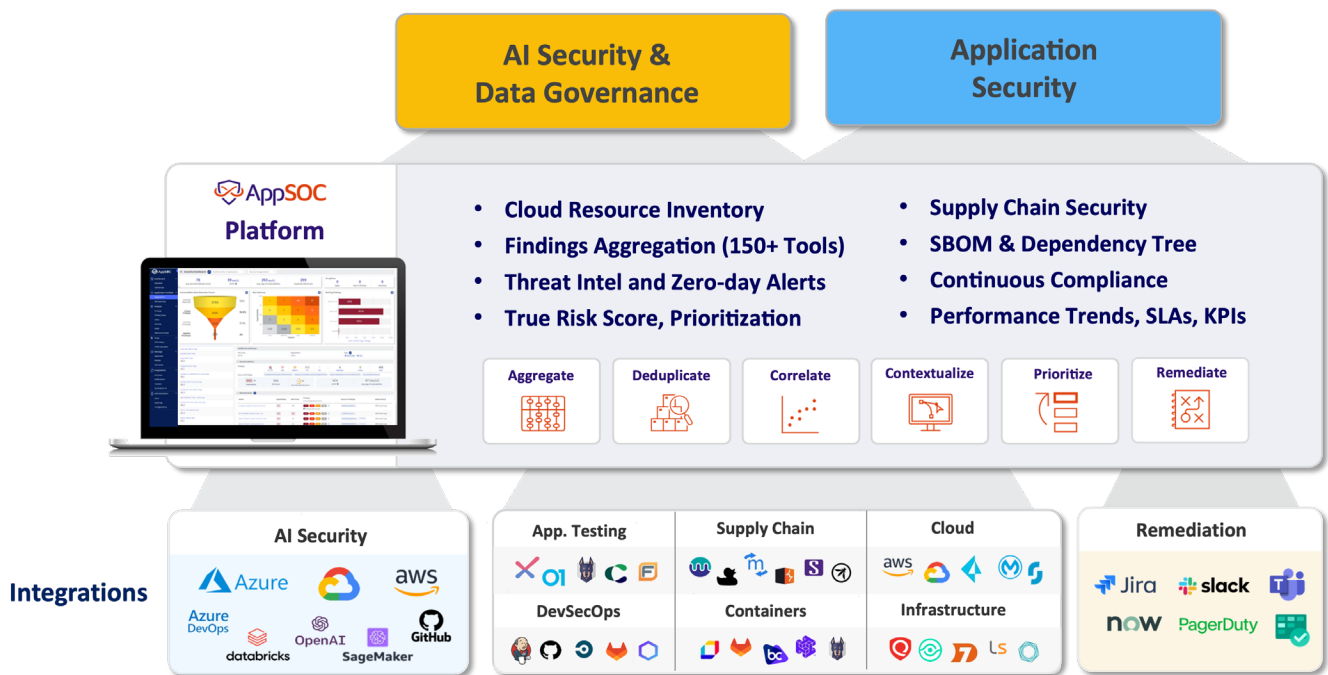


# AppSOC AI Application Security & Governance

## Extending the AppSOC Platform

AppSOC 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
<ul style="list-style-type: none"> <li>Shadow AI discovery</li> <li>AI use-case repository</li> <li>Ownership &amp; policy mapping</li> <li>AI impact assessments</li> </ul>	<ul style="list-style-type: none"> <li>AI stack visibility (models, integrations, consumers)</li> <li>Resource inventory</li> <li>Configuration hardening</li> <li>KB &amp; risk rating</li> </ul>	<ul style="list-style-type: none"> <li>Operational risk scanning (security &amp; licensing)</li> <li>Model scanning</li> <li>Prompt injection detection</li> <li>Runtime policy enforcement (bias, acceptable use)</li> </ul>	<ul style="list-style-type: none"> <li>Data classification &amp; lineage</li> <li>DLP (PCI/PHI/PII)</li> <li>Data access anomalies</li> <li>Identity &amp; application mapping</li> </ul>



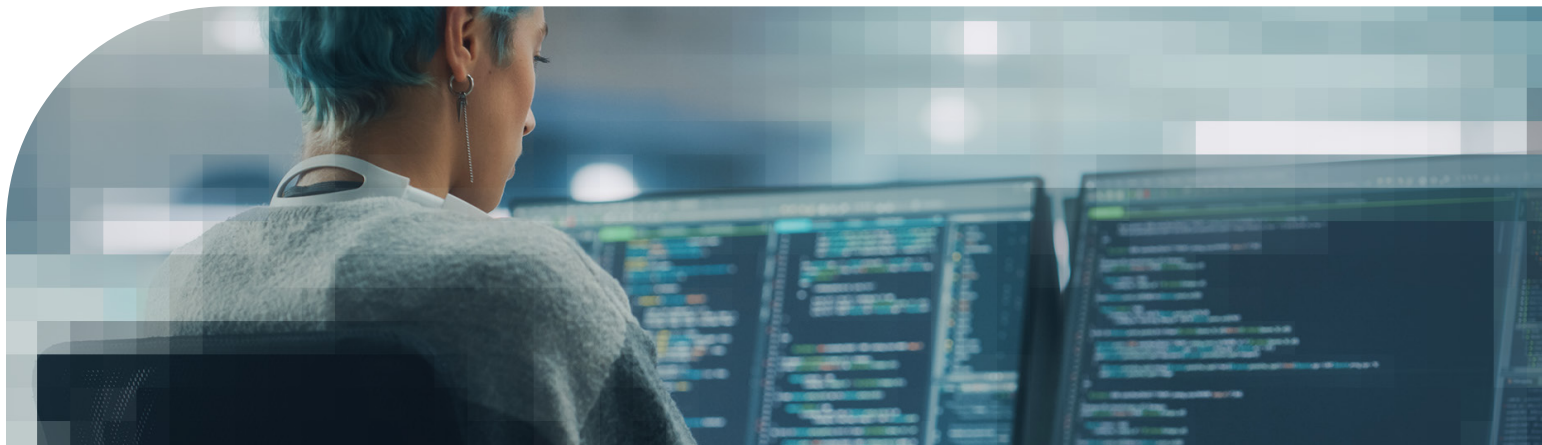
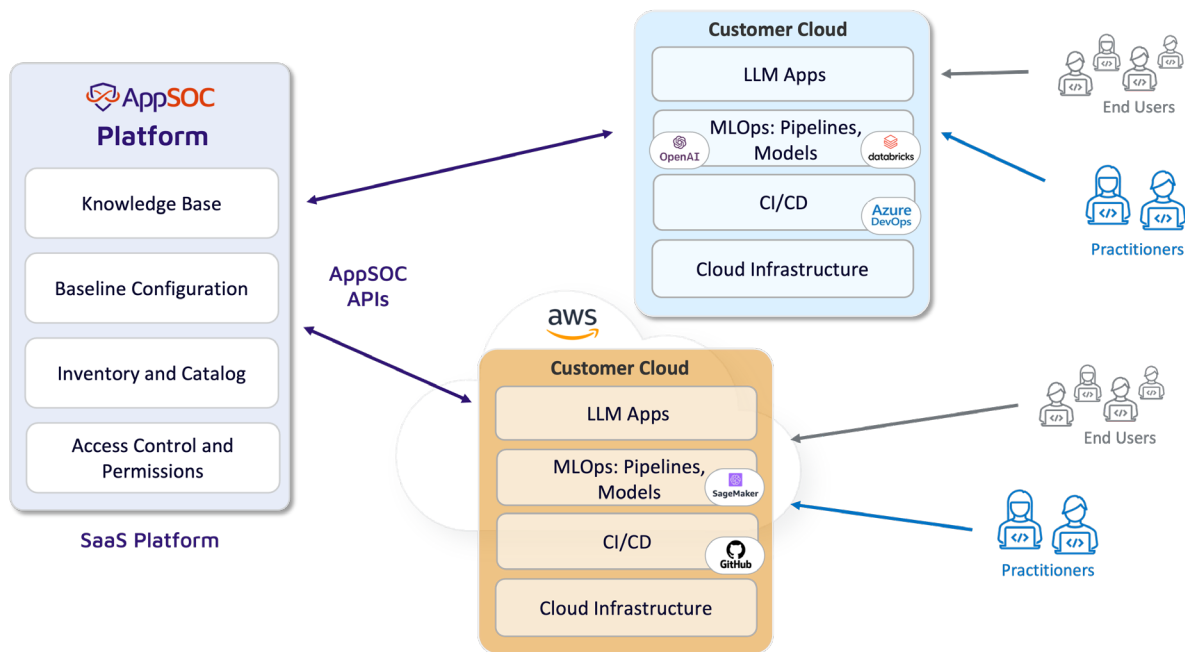
## Frictionless Deployment

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



AppSOC 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. AppSOC's global team is headquartered in California in the heart of Silicon Valley.

Learn more at [AppSOC.com](https://www.AppSOC.com)

Follow us on 

All other names mentioned herein are trademarks or registered trademarks of their respective owners. AppSOC-Platform-SB-041723