

SPEAKERS



PARTNER

ANA BRUDER

MAYER BROWN



STEPHEN LILLEY

MAYER BROWN



ASSOCIATE VP, ASSISTANT
GENERAL COUNSEL - CYBER

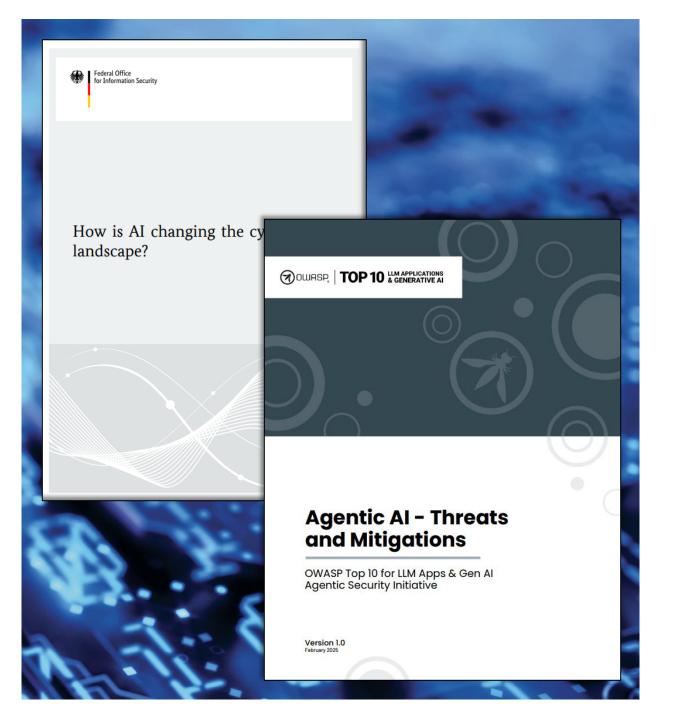
JIM SFEKAS

FILLILLY

AGENDA

- 1. Threats to Al
- 2. Legal Expectations for Securing Al
- 3. Implementing a Risk-Based AI Security Program





AI SECURITY THREATS

- Companies face a broad range of attacks on AI systems, including attacks that are common to other software-based systems and attacks that are distinctive to AI systems. Attacks include:
 - Evasion attacks: malicious input to fool the model or reduce its accuracy,
 e.g., prompt injection
 - Poisoning attacks, e.g., data poisoning, model poisoning
 - Information extraction attacks, e.g., model stealing, data reconstruction, membership or attribute inference attacks
 - Supply chain attacks, e.g., slopsquatting
 - Abuse of agentic Al
- Companies also face inadvertent security risks from the use of Al, including from the use of shadow Al or the use of sensitive data in model finetuning or prompts
- Companies can turn to an increasing number of resources to understand these threats, such as NIST, OWASP, MITRE Atlas, German BSI

02 LEGAL EXPECTATIONS FOR SECURING AI

AI SECURITY BEST PRACTICES WILL INFORM LEGAL EXPECTATIONS FOR COMPANIES

- Best practices for AI security have been developed along a number of key dimensions of AI security, including:
 - Data security
 - Application security
 - Model/model weight security
 - Infrastructure security
 - Securing AI output (code development)
- Companies also face continued—and potentially heightened expectations to maintain appropriate security for the IT on which Al systems depend.
- How exactly these best practices will inform regulatory expectations, litigation claims, and contractual requirements remains to be seen.

Joint Cybersecurity Information

TLP:CLEAR















Al Data Security

Best Practices for Securing Data Used to Train & Operate AI Systems

Executive summary

This Cybersecurity Information Sheet (CSI) provides essential guidance on securing data used in artificial intelligence (AI) and machine learning (ML) systems. It also highlights the importance of data security in ensuring the accuracy and integrity of AI outcomes and outlines potential risks arising from data integrity issues in various stages of AI development and deployment.

This CSI provides a brief overview of the AI system lifecycle and general best practices to secure data used during the development, testing, and operation of AI-based systems. These best practices include the incorporation of techniques such as data encryption, digital signatures, data provenance tracking, secure storage, and trust infrastructure. This CSI also provides an in-depth examination of three significant areas of data security risks in AI systems: data supply chain, maliciously modified ("poisoned") data, and data drift. Each section provides a detailed description of the risks and the corresponding best practices to mitigate those risks.

This guidance is intended primarily for organizations using AI systems in their operations, with a focus on protecting sensitive, proprietary, or mission critical data. The principles outlined in this information sheet provide a robust foundation for securing AI data and ensuring the reliability and accuracy of AI-driven outcomes.

This document was authored by the National Security Agency's Artificial Intelligence Security Center (AISC), the Cybersecurity and Infrastructure Security Agency (CISA), the Federal Bureau of Investigation (FBI), the Australian Signals Directorate's Australian Cyber Security Centre (ASD's ACSC), the New Zealand's Government Communications

This information is marked TLP:CLEAR. TLP:CLEAR information may be distributed without restriction. For more information on the Traffic Light Protocol, see cisa.gov/tlp/.

U/OO/157249-25 | PP-25-2301 | May 2025 Ver. 1.0



LEGAL RISKS ARE SIGNIFICANT DESPITE LIMITED SPECIFIC LEGAL REQUIREMENTS

- The EU AI Act provides limited guidance on security expectations:
 - Security at system level, but taking into the account other dimensions
 - Guiding principles:
 - Compliance at a system level
 - Security risk assessments needed
 - Integrated and continuous approach
 - Limits in the state of the art for securing Al models
- However, there are numerous legal frameworks, including many that are sector specific, that inform legal expectations for AI security.



03

IMPLEMENTING A RISK-BASED AI SECURITY PROGRAM

Guidelines for secure Al system development



Joint Cybersecurity Information









Deploying AI Systems Securely

Best Practices for Deploying Secure and Resilient Al Systems

Executive summary

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

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 Al system development and incorporates mitigation considerations from Engaging with Artificial Intelligence (AI). It is for organizations deploying and operating AI systems designed and developed by another entity. The best practices may not be applicable to all environments, so the mitigations should be adapted to specific use cases and threat profiles. [1], [2]

Al security is a rapidly evolving area of research. As agencies, industry, and academia discover potential weaknesses in AI technology and techniques to exploit them, organizations will need to update their Al systems to address the changing risks, in addition to applying traditional IT best practices to AI systems.

This report was authored by the U.S. National Security Agency's Artificial Intelligence Security Center (AISC), the Cybersecurity and Infrastructure Security Agency (CISA), the Federal Bureau of Investigation (FBI), the Australian Signals Directorate's Australian Cyber Security Centre (ACSC), the Canadian Centre for Cyber Security (CCCS), the New Zealand National Cyber Security Centre (NCSC-NZ), and the United Kingdom's National Cyber Security Centre (NCSC-UK). The goals of the AISC and the report are

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

subject to standard copyright rules. For more on the Traffic Light Protocol, see cisa gov/flo



IMPLEMENTING A RISK-BASED AI SECURITY PROGRAM WILL HELP A COMPANY CAPTURE THE BENEFITS OF AI ADOPTION

General cyber risk measures

- Threat modeling, risk assessment, and vulnerability testing
- Strong access controls, identity management, and permission management (e.g. principle of least privilege)
- Supply chain security and component provenance
- Logging, monitoring, and incident response planning

Al-specific measures

- Data provenance, integrity, and bias assessment for training data
- Adversarial testing, red teaming, and guardrails for prompt injection
- Monitoring for model drift, data poisoning, and misuse
- Documentation of model limitations, intended use, and failure modes

Key areas to consider when implementing an Al security program include:

- Policies and controls Governance
 - Security testing Procurement

EFFECTIVE GOVERNANCE CAN REDUCE RISKS ASSOCIATED WITH AI SECURITY

- Poor calibration of AI security can have significant consequences for a company, whether because it prevents the company from innovating at the necessary pace or because it exposes the company to excessive risk that undermines the benefits of that innovation
- The security team will be an important voice in determining how to manage Al security risk, but this issue will also implicate the expertise and interest of relevant business units, legal, and other stakeholders.
- As in other AI contexts, an effective governance mechanism will help the company appropriately manage AI security risks. This governance will be most effective if it:
 - Includes appropriate stakeholders;
 - Is informed by appropriate risk assessments;
 - Has full visibility into AI deployments across the company;
 - Has authority to impose necessary security measures and processes;
 - Can guide investment decisions into Al-specific security tools;
 - Is implemented through appropriate policies, controls, and procedures;
 - Allows effective executive oversight and decision-making of Al security.

KEY QUESTIONS

- Are necessary stakeholders engaged in managing Al security?
- Does Al security governance fit with other governance mechanisms (e.g., Al, security more broadly)?
- Does Al security governance reach from technical controls to executive decision-making?

RED FLAGS

- Security team is not included in Al governance mechanism
- Development team can disregard security considerations.

FOCUSING ON SECURITY IN AI PROCUREMENT CAN SUBSTANTIALLY REDUCE RISK

 The procurement process can highlight the potential tension between innovation through rapid adoption of AI tools and ensuring appropriate security that allows the company to fully benefit from that innovation.

Focus on third-party risk

- Heightened emphasis on third party risk in recent cyber regulations
- Particularly relevant in AI context: many layers in supply chain

Considerations for procurement teams and their counsel

- Take time to understand product and security risk
- Include AI-specific questions on vendor questionnaires
- Assess need for security-specific terms to address AI security in contracts with Al vendors
- Consider impact on other terms, like breach notification, liability

KEY QUESTIONS

- What level of risk does the service provided by the vendor present to the organization?
- Does the vendor meet prevailing security best practices relevant to the service it provides?
- Will the vendor agree to security provisions appropriate to the risk presented by its service?

RFD FLAGS

- Vendor lacks appropriate security maturity.
- Scope of service is unclear or could expand over time.

APPROPRIATE POLICIES AND CONTROLS CAN REDUCE AI SECURITY RISK

- Security policies and controls are likely to vary based on the nature of the company's business and the AI use case, including the sensitivity of the data it will access and the scope of actions it can trigger/take.
- As a baseline, the security policies and controls that apply to other softwarebased systems presumptively should apply to AI systems to the extent feasible.
- Key issues for attention include: (1) Al tool permissions, for data access and permitted actions; (2) user access rights; (3) system logging and monitoring;
 (4) data loss prevention; and (5) integration of security into Al development activities.
- With Al-specific security solutions proliferating in the market, security controls should be increasingly automated and security practices should avoid over-reliance on guidelines for user behavior.
- Companies may wish to update their security policies to address the use of Al or to create specific Al security policies or processes.

KEY QUESTIONS

- Are security policies and controls based on an appropriate assessment of relevant risks?
- Are Al systems built on a weak foundation in that relevant infrastructure lacks appropriate controls?
- Do security controls leverage available technological solutions in an effective way?

RED FLAGS

- Al is implemented with a deployfirst, secure-later mindset
- Paper security policies do not match implemented controls

TESTING THE SECURITY OF AI SYSTEMS WILL HELP ONGOING RISK MITIGATION ACTIVITIES

- In addition to more traditional security testing, Al red-teaming has important distinctive elements:
 - Involves adversarial testing methods, e.g., attempts to elicit unwanted behaviors, subvert the model's built-in defenses or quardrails
 - Context-dependent: Red-teaming practices and objectives vary by stakeholder (e.g., commercial developers vs. national security organizations) and by model type (general-purpose vs. specialized models)

Challenges:

- Measurement: what does it mean to "break" a model, and what constitutes a model failure or vulnerability?
- Testing across multiple models and tracking results over time
- Building consensus around testing practices and maintaining transparency
- Particular questions for frontier models

KEY QUESTIONS

- Does Al red-teaming account for the distinctive risks associated with Al systems?
- Should the testing be performed at the direction of counsel and the reports subject to legal privilege?
- Are test results incorporated into relevant risk assessments so that they can prioritized along with other key risks?

RFD FLAGS

- Red-teaming is not tailored to the particular circumstances
- Red-teaming does not inform decision making in a practical way

Questions? THANK YOU!

MAYER | BROWN

This Mayer Brown publication provides information and comments on legal issues and developments of interest to our clients and friends. The foregoing is not a comprehensive treatment of the subject matter covered and is not intended to provide legal advice. Readers should seek legal advice before taking any action with respect to the matters discussed herein.

Mayer Brown is a global legal services provider comprising associated legal practices that are separate entities, including Mayer Brown LLP (Illinois, USA), Mayer Brown International LLP (England & Wales), Mayer Brown Hong Kong LLP (a Hong Kong limited liability partnership) and Tauil & Chequer Advogados (a Brazilian law partnership) (collectively, the "Mayer Brown Practices"). The Mayer Brown Practices are established in various jurisdictions and may be a legal person or a partnership. PK Wong & Nair LLC ("PKWN") is the constituent Singapore law practice of our licensed joint law venture in Singapore, Mayer Brown PK Wong & Nair Pte. Ltd. More information about the individual Mayer Brown Practices and PKWN can be found in the Legal Notices section of our website.

"Mayer Brown" and the Mayer Brown logo are the trademarks of Mayer Brown. © 2025 Mayer Brown. All rights reserved.