INSIGHTS

Privacy & Security

NSA Issues Cybersecurity Guidance and Best Practices for AI Systems

The recommendations—and NSA's collaboration with other agencies on the topic—indicate the growing importance of securing AI systems and data

By Michael T. Borgia and Andrew M. Lewis

06.11.25

The National Security Agency (NSA), in coordination with the Cybersecurity and Infrastructure Security Agency (CISA), the Federal Bureau of Investigation (FBI) and cybersecurity agencies from Australia,

New Zealand, and the United Kingdom, has released a Cybersecurity Information Sheet (CSI) providing recommendations to address a growing national and international priority: ensuring that the data used to develop, operate, and maintain AI models remains secure, trustworthy, and resistant to tampering. The CSI, titled *AI Data Security: Best Practices for Securing Data Used to Train & Operate AI Systems*, was released on May 22, 2025. The CSI builds on the NSA's joint guidance on Deploying AI Systems Securely, which was issued in April 2024 in coordination with CISA, the FBI, and multiple international partners.

The CSI identifies three key risks to the performance, accuracy, and integrity of AI and ML systems, and provides 10 recommendations primarily for civilian and defense agencies, as well as private-sector operators of critical infrastructure.

Key Risks to Data Used in AI Systems

The CSI outlines three overarching categories of data security threats that span the AI lifecycle:

1. **Data Supply Chain Risks:** AI systems rely on external, large-scale datasets. Those datasets may be crowdsourced or carefully curated. Regardless of how the datasets are sourced, they are vulnerable to manipulation and degradation by untrusted third parties. Compromised data upstream—such as web-crawled images, expired domains, or poisoned media files—can propagate through AI training pipelines, influencing future outputs and system behavior.

2. **Maliciously Modified ("Poisoned") Data:** Threat actors may intentionally insert adversarial or false data into training sets to manipulate model behavior. This type of "data poisoning" includes both obvious threats like inserting intentional disinformation and

more subtle ones like introducing statistical bias or manipulating metadata. Poisoned data may result in unsafe AI behavior or performance degradation, particularly in high-stakes domains such as cybersecurity or defense targeting. Where AI is used to identify and stop cyber threats, data poisoning can be used to make cyber attacks more difficult to detect.

3. **Data Drift:** Once deployed, AI systems face the risk of "data drift"—a gradual or sudden shift in the statistical properties of incoming data compared to what the model was originally trained on. Drift can degrade system accuracy over time, obscure early warnings of performance failures, or be exploited by malicious actors to bypass AI-driven safeguards. Unlike poisoning, which is deliberate, drift may occur naturally through environmental or operational change.

The CSI aligns these three risk areas to each of the major stages in the lifecycle of AI systems, as identified in the AI Risk Management Framework published by the National Institute of Standards and Technology (NIST). According to the CSI, addressing these three key risk areas requires a combination of technical controls, operational oversight, and continuous vigilance throughout the AI system lifecycle. The CSI addresses those activities in its 10 recommendations for securing AI data.

Best Practices To Secure Data Used in AI Systems

The CSI recommends 10 sets of practices and a variety of specific controls for securing data used in AI and ML systems. These practices and controls are not specific to AI—they may be used to secure sensitive data in a variety of contexts—but according to the CSI are particularly well-suited to address the key risks to AI systems and data described above.

We summarize each of the 10 recommendations and explain their relevance below.

1. Source Reliable Data and Track Data Provenance

Data provenance (*i.e.*, the history and origin of data) is foundational for establishing the trustworthiness and reliability of AI systems. Without an established provenance, there may be significant uncertainty as to whether the data used to train or operate AI systems is accurate, current, and uncompromised. Web-scale datasets—*i.e.*, very large datasets derived from Internet sources, typically through web scraping—are especially vulnerable to data poisoning, disinformation, or outdated content.

The CSI recommends:

- Only ingesting data from trusted, reliable and—to the extent feasible—authoritative sources.
- Maintaining detailed records of how the data was obtained, processed, and modified.
- Using cryptographically signed, append-only ledgers to track changes to data.
- Adding digital "content credentials" to media and documents to support verifiability and authenticity. Content credentials use cryptography to bind metadata to media content, allowing users to track the file's origins and modifications.
- Requiring formal certifications from dataset and model providers attesting that their data is free from compromise.

2. Verify and Maintain Data Integrity During Storage and Transport

Even if an organization sources reliable, trustworthy data, that data can be corrupted during transit or storage. Data corruption risks misleading the AI system and undermining its predictions or decisions.

The CSI recommends use of cryptographic hashes and checksums to

verify that data has not been altered. These tools can detect even subtle changes to files and signal that the data should be quarantined or discarded. The CSI also recommends that organizations conduct periodic audits and testing of training data to remove inaccurate information and identify other quality issues.

3. Employ Digital Signatures To Authenticate Trusted Data Revisions

Datasets used in AI and ML systems will be updated and adjusted over time, including through model training, fine tuning, and real-world feedback. Organizations must adopt strong protections to authenticate these revisions and confirm that they are from trusted sources.

The CSI recommends that organizations digitally sign all original datasets and subsequent revisions to prevent tampering and identify illegitimate changes to datasets. The CSI further recommends that organizations adopt quantum-resistant cryptographic standards to future-proof authentication methods in light of growing threats to cryptography from quantum computing.

4. Leverage Trusted Infrastructure

If attackers compromise the infrastructure used to process or store training data, they may not need to poison the data to create malicious outcomes. Attackers with access to the AI system's underlying infrastructure may be able to simply alter how the system interprets the data.

The CSI recommends that organizations employ "Zero Trust" security architecture principles, such as by creating secure enclaves—*i.e.*, hardware-based, isolated areas of a computer processor—to protect sensitive operations. Moving to a Zero Trust security model has been a major priority for the federal government over the last several years, as we have discussed in prior posts.

5. Classify Data and Use Access Controls

Data classification is a common security technique that involves classifying data according to its sensitivity, business impact, or other factors, and applying a standard set of security controls to the data based on its classification. The CSI recommends that organizations classify both inputs and outputs of AI systems and classify a system's inputs and outputs at the same classification level.

6. Encrypt Data

Encryption safeguards the confidentiality and integrity of data, even if the system on which the data is processed has been compromised. The CSI recommends that encryption be applied at every stage of data processing for AI systems: at rest, in transit, and during computation. The CSI further recommends that organizations use industry standard, quantum-resistant encryption algorithms such as AES-256.

The CSI recommends use of FIPS 140-3 compliant encryption algorithms, with AES-256 for standard security needs and post-quantum algorithms where applicable, and employ TLS 1.3 for data-in-transit.

7. Store Data Securely

Secure storage may be the last line of defense against data exfiltration or corruption once a system has been compromised. The CSI recommends use of certified storage systems with access logging, tamper detection, and cryptographic tools that comply with FIPS 140-3.

8. Leverage Privacy-Preserving Techniques

AI systems process large amounts of personal data, much of which may be sensitive. The CSI recommends that organizations adopt various privacy preserving technologies, including:

- **Data masking** to anonymize inputs during training by replacing real

personal information with fake but realistic training data.

- **Differential privacy** techniques when releasing aggregate results or training on personal data, to minimize the risk that an individual could be identified from that aggregate data.
- **Federated learning** and **secure multi-party computation**, both of which involve processing separate datasets to preserve the privacy of individuals whose data has been included in one of those sets.

9. Delete Data Securely

Retention of data that is no longer needed to support an AI system creates excess liability risk. And if such data is not securely deleted, attackers may be able to restore that data and exploit it.

The CSI recommends use of NIST SP 800-88-compliant data sanitization methods for media used for AI data storage and processing. Those sanitization methods include cryptographic erase and data overwriting.

10. Conduct Ongoing Data Security Risk Assessments

Threats to AI data evolve as quickly as the technology itself. Systems that are secure against today's threats may be insecure against emerging threats. The CSI recommends that organizations conduct regular risk assessments of their AI systems using frameworks like NIST's Risk Management Framework (RMF) and AI Risk Management Framework to identify and prioritize risks across the AI data lifecycle.

Conclusion

Organizations that deploy AI and ML systems should carefully review and consider the guidance set forth in the CSI. Among other things, it is important for such organizations to assess how they can mitigate emerging and constantly shifting risks to AI systems and data, including risks posed by quantum computing and novel attack strategies against AI

datasets and infrastructure.

DWT's Privacy & Security and Artificial Intelligence teams will continue to monitor updates in this area.

Related Insights

10.02.25
INSIGHTS

CISA 2015 Has Sunset. Now What?

09.17.25
INSIGHTS

CISA Delays Cyber Incident Reporting Rules Until May 2026

09.12.25
INSIGHTS

Department of Defense Issues Final Rule to Implement Cybersecurity Maturity Model Certification (CMMC) Program