NIST Releases Draft Cybersecurity Framework Profile for AI

Date: 2025-12-17

The National Institute of Standards and Technology (NIST) released an Initial Preliminary Draft of *NIST IR 8596: Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile)*. The document is a CSF 2.0 Community Profile designed to help organizations manage cybersecurity risks associated with artificial intelligence systems. https://nvlpubs.nist.gov/nistpubs/ir/2025/NIST.IR.8596.iprd.pdf

The Cyber AI Profile is organized around the NIST Cybersecurity Framework 2.0 Functions, Categories, and Subcategories. It does not replace existing frameworks, instead it prioritizes cybersecurity outcomes specifically for AI-related risks.

The profile addresses three AI-focused cybersecurity areas:

- Securing AI System Components (Secure), covering models, data, infrastructure, and supply chains
- Conducting AI-Enabled Cyber Defense (Defend), addressing the use of AI to enhance cybersecurity operations
- Thwarting AI-Enabled Cyber Attacks (Thwart), focusing on resilience against adversarial uses of AI.

The draft explicitly states that AI introduces expanded and altered attack surfaces, including risks related to training data, model behavior, supply chains, and system integration. It also notes that AI behavior can be opaque, dynamic, and difficult to predict, making security controls insufficient on their own.

The document emphasizes that organizations must integrate AI considerations into existing governance, risk management, legal compliance, and cybersecurity programs, rather than treating AI as a separate technical issue.

The draft is open for public comment from December 16, 2025 through January 30, 2026.

The draft does not yet provide finalized mappings to specific regulatory requirements, such as sector-specific compliance obligations. While it references standards like NIST SP 800-53 and the AI Risk Management Framework, the practical enforcement expectations for regulated entities remain undefined.

The guidance is intentionally technology-neutral, a design choice that enables broad applicability but leaves open questions around how organizations operationalize controls for specific AI use cases such as generative AI, agentic systems, and third-party AI services.

These gaps are a central focus of the pubic comment process and are expected to be refined in subsequent versions.

The documents also acknowledges that many AI-enabled attacks may go undetected today, but does not define measurable thresholds for acceptable AI-related risk or minimum control baselines.

While the draft guidance outlines high-level principles for securing AI systems, it leaves open critical questions about how organizations will translate these expectations into governance, documentation, and oversight practices. For legal, compliance, and risk teams, the key challenge is not whether AI should be secured, but how AI-related risks are defined, documented, audited, and enforced within existing enterprise frameworks.

Core questions for organizations

- How will organizations demonstrate compliance with AI-related cybersecurity expectations during internal reviews, external audits, or regulatory examinations?
- How should AI-specific risks be documented within existing enterprise risk registers, including risk ownership, severity, and mitigation tracking?
- What governance structures are required to oversee AI security decisions across legal, technical, compliance, and business teams, and who retains final accountability.
- How should organizations assess, monitor, and document third-party AI vendors, model providers, and AI supply chains under this framework?
- At what point does AI system behavior, such as data leakage, model manipulation, or autonomous decision errors, trigger incident response procedures, or regulatory reporting obligations?

This draft reflects a clear shift on how AI cybersecurity risk is being framed. Rather than treating AI security as a purely technical concern, the guidance positions it as a governance, risk, and compliance issue that must be integrated into existing cybersecurity management structures.

Organizations deploying AI systems are implicitly expected to incorporate AI into their broader cybersecurity risk management strategies, rather than managing it as a standalone initiative. This includes reorganizing and addressing legal, regulatory, and contractual obligations associated with AI use, as well as preparing for audits and assessments in which AI system security is evaluated alongside traditional IT controls. The draft also signals that expectations around accountability, transparency, and oversight of AI systems will continue to evolve, increasing regulatory scrutiny over time.

For compliance, governance, and risk professionals, this draft provides an early indication of how AI security may be assessed under existing frameworks such as NIST CSF and enterprise risk management programs. Organizations that begin aligning AI governance, documentation, and oversight practices now are likely to be better positions to respond to future regulatory guidance, enforcement actions, and audit requirements.

Sources

- Primary documents: NIST IR 8596 (Initial Preliminary Draft), *Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile)*, December 2025 https://nvlpubs.nist.gov/nistpubs/ir/2025/NIST.IR.8596.iprd.pdf
- Public comment period: December 16, 2025-January 30, 2026