KPMG

Regulatory Alert

Regulatory Insights

February 2026

Cybersecurity: NIST Draft Cybersecurity Framework for AI

KPMG Regulatory Insights:

- ***New Cyber AI Profile:*** *Extends the Cybersecurity Framework to new cyber risks introduced by AI; initial preliminary draft of this Cyber AI Profile will inform future proposals.*
- ***Layering:*** *Existing cybersecurity and AI guidance would remain in place; AI-specific priorities would be layered onto the CSF 2.0.*
- ***Governance:*** *Recognizes AI as a cybersecurity governance concern.*
- ***Benchmarks:*** *Though voluntary, the Cyber AI Profile may potentially serve as a benchmark for regulators and others regarding cybersecurity diligence.*

The U.S. Department of Commerce's National Institute of Standards and Technology (NIST) has released an initial preliminary draft of the *Cybersecurity Framework Profile for Artificial Intelligence* (Cyber AI Profile or NIST IR 8596*)*. The preliminary draft is designed as a voluntary framework that would extend the recently updated NIST Cybersecurity Framework (CSF) 2.0 to new cybersecurity risks and opportunities introduced by AI and to also complement NIST's AI Risk Management Framework (AI RMF). NIST envisions the CSF 2.0, the AI RMF, and the Cyber AI Profile being used together.

The preliminary draft of the Cyber AI Profile is organized around:

- Three Focus Areas: Secure (securing AI systems); Defend (conducting AI-enabled cyber defense); and Thwart (thwarting adversarial cyberattacks using AI).
- Six CSF 2.0 Core Functions: Govern, Identify, Protect, Detect, Respond, and Recover.

In a separate but related release, NIST also made available a discussion draft covering "*Control Overlays for Securing AI Systems*" including "*Overview and Methodology*" (NIST IR 8605) and "*Using and Fine-Tuning Predictive AI*" (NIST IR 8605A), which will serve as complements to the Cyber AI Profile.

Overview of NIST Cyber AI Profile (Preliminary Draft)

Focus Areas:

The core concept of the Cyber AI Profile would be to apply the structure of the CSF 2.0 and the AI RMF to AI specific risks rather than to create a new, separate framework. As presented in the initial preliminary draft, it would provide guidelines for managing cybersecurity risk related to AI systems as well as identifying opportunities for using AI to enhance cybersecurity capabilities and leverage AI as a defensive tool. In this context, AI is meant to refer to any systems that are using AI capabilities, whether they are stand-alone AI systems or applications, or infrastructure including LLMs, predictive analytics, generative AI, agentic AI, and search engines.

The draft proposes a risk-based approach organized around three key focus areas:

- **Secure:** Securing AI systems by concentrating on managing cybersecurity challenges when integrating AI into organizational ecosystems and infrastructure. It would include AI systems, their supply chains, data and machine learning infrastructure, and the other systems and data on which AI relies.
- **Defend:** Conducting AI-enabled cyber defense by identifying opportunities to use AI to enhance cybersecurity processes and activities. It would include areas such as "mission assurance" (e.g., security

governance and policy), "predictive and proactive" (e.g., identifying potential threat actors and tactics), "investigation and analysis" (e.g. advanced threat detection and analysis, zero trust modeling), and "response and remediation" (e.g., adversarial training and simulation, incident reporting).

— **Thwart:** Thwarting AI-enabled cyberattacks focus by building resilience to protect against new AI-enabled threat vectors. It would include how AI can advance capabilities of adversaries, the impact of those attacks, and what organizations can do to bolster their systems against emerging threats.

Functions:

The preliminary draft proposes to integrate AI-specific considerations across all six core functions of the NIST CSF 2.0. Sample considerations are provided for each of the three focus areas. The considerations are assigned a proposed priority level - "1" for High Priority, "2" for Moderate Priority, and "3" for Foundational Priority - to convey the areas to address sooner and to guide planning to reach the intended cybersecurity outcome. However, the priority levels may be higher or lower for individual organizations based on characteristics of the environment, needs, risk tolerance, or other factors. Organizations are expected to make the decision to deploy AI-related cybersecurity mitigations based on their own unique needs and risk tolerances. The core functions include:

— **Govern (GV):** The goal for this function is for the organization's cybersecurity risk management strategy, expectations and policy to be established, communicated, and monitored. The categories covered include organizational context; risk management strategy; roles, responsibilities and authorities; policy; oversight; and cybersecurity supply chain risk management. Proposed AI considerations include:

- Communicating the intended use and known limitations of AI.
- Identifying business outcomes that rely on AI systems and communicating dependencies to relevant teams.
- Implementing continuous monitoring and threat detection across supplier-provided AI models, datasets, and APIs to identify adversarial behaviors, data leakage, or compromised components originating from the supplier.

— **Identify (ID):** The expected outcome for this function is for the organization's cybersecurity risks to be understood. The covered categories include asset management; risk assessment; and improvement (to processes, procedures and activities). Proposed AI considerations include:

- Identifying, tracking, and recording new classes of vulnerabilities from AI.
- Understanding the nature of the data and metadata, as well as the requirements that travel with them (e.g., use agreements, consent).
- Including AI-specific procedures for containment (e.g., disabling model autonomy), triage (e.g., analyzing model logs), and recovery (e.g., restoring validated model versions) in incident response plans.

— **Protect (PR):** This function outlines safeguards to manage the organization's cybersecurity risks. It covers identity management, authentication, and access control; awareness and training; data security; platform security; and technology infrastructure resilience. Proposed AI considerations include:

- Creating separate identities and credentials for AI systems (i.e., AI service level accounts) that interact with broader systems including AI defense agents to support defensive response activities.
- Establishing new policies to govern the permissions/access controls and authorizations for AI systems.
- Maintaining strict guidance on downloading and installing software into production systems.
- Providing personnel with access to training/information about new, emerging, AI-enabled threats.

— **Detect (DE):** Sets out approaches to detect and analyze cybersecurity attacks and compromises. The categories covered include continuous monitoring (of assets for potentially adverse events) and adverse event analysis. Proposed AI considerations include:

- Establishing new monitoring procedures to track actions taken by AI such as to detect adversarial inputs or anomalous AI system behaviors.
- Understanding the potential scope and scale of AI-enabled cyber attacks.
- Aggregating data from multiple log sources to enhance AI-cyber defenses in detecting anomalous and potentially adverse events.

— **Respond (RS):** The Respond function anticipates that actions will be taken to address a detected cybersecurity incident and includes categories for incident management; incident analysis; incident response reporting and communication; and incident mitigation. Proposed AI considerations include:

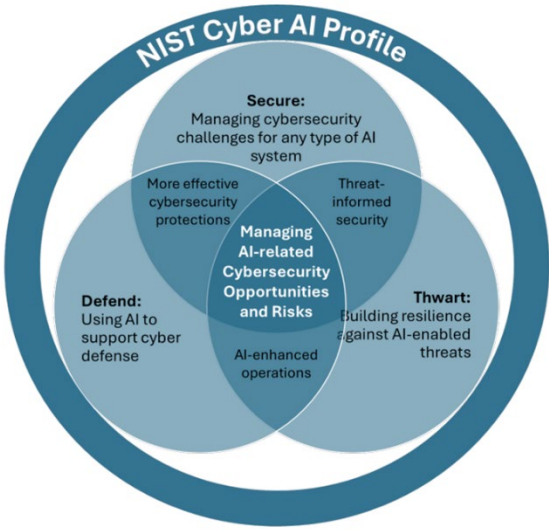
- Employing new expertise, tools, and methods to diagnosis attacks on AI such as explicitly conducting searches for indicators of adversary AI usage in incident analysis.
- Preserving logs, inputs, outputs, and decision chains of AI systems to ensure provenance and improving future AI-driven response actions.
- Tracking datasets and versioning as well as documentation of associated metadata related to the model.
- Implementing automated actions (e.g., blocking, isolating systems) in addition to AI-enabled defenses to flag adverse behaviors and events for review.

— **Recover (RC):** The Recover function supports the restoration of assets and operations affected by a cybersecurity incident. The categories that would be covered include incident recovery plan execution and incident recovery communication. Proposed AI considerations include:

- Testing model and dataset backups for poisoning or drift to assure the integrity of AI components.
- Using AI to evaluate which systems to restore first, track progress, and draft updates to stakeholders.
- Verifying the integrity of restored AI components (models, training data) for compromise (e.g., residual poisoning) and validating that the restored AI defense system operates at expected performance (e.g., model accuracy, FP rate) before confirming normal operational status.

For more information, please contact Bryan McGowan, Katie Boswell or Laura Byerly.

Relationship Between Cyber AI Profile Focus Areas



Source: NIST IR 8596, Cybersecurity Framework Profile for Artificial Intelligence

Contact

Laura Byerly
Managing Director
Regulatory Insights
lbyerly@kpmg.com

Brian Hart
Principal
Risk, Regulatory and Compliance
bhart@kpmg.com