Cyber AI Profile Project Overview

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Cybersecurity, Privacy, and AI

The diverse use and rapid proliferation of Artificial Intelligence (AI) promises unique value for industry, consumers, and broader society, but like many technologies, to recognize these benefits to the greatest potential, new risks from these advancements in AI must be managed.

In NIST's Applied Cybersecurity Division (ACD), our key concern is how advancements in the broad adoption of AI may impact current cybersecurity and privacy risks and risk management approaches.

https://www.nist.gov/itl/applied-cybersecurity/cybersecurity-privacy-and-ai

NIST AI and Cybersecurity Projects

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Topic	Learn More!
AI Risk Management Framework (AI RMF) A framework to better manage risks to individuals, organizations, and society associated with artificial intelligence	
Center for AI Standards and Innovation (CAISI) Facilitates testing and collaborative research related to harnessing and securing the potential of commercial AI systems	
Adversarial Machine Learning Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations (NIST AI 100-2 E2025)	
Dioptra A software test platform for assessing the trustworthy characteristics of artificial intelligence systems	
Secure Software Development Framework (SSDF) AI Profile Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile	

Topic	Learn More!
PETs Test Bed Evaluating Differential Privacy Guarantees	
DevSecOps Secure Software Development, Security, and Operations (DevSecOps) Practices	
Agent Identities Digital Identity Guidelines, Revision 4 (NIST SP 800-63)	
NCCoE Chatbot Secure, internal-use chatbot to assist with discovering and summarizing cybersecurity guidelines	
COSAiS NIST SP 800-53 Control Overlays for Securing AI Systems (COSAiS)	

What We Heard

- CISO's are concerned with how to strategically address cybersecurity as a result of advancements in AI but their hands are already full dealing with current ongoing operations and they could benefit from prioritization
- There is already much ongoing discussion and work in many of these areas but there is no consistent taxonomy or relation to an organizations strategic cybersecurity risk management
- There are some new or modified impacts to cybersecurity but do not reinvent the wheel, rather build on existing frameworks or cybersecurity practices and identify the what is new
- There is limited overlap between cybersecurity practitioner/training and AI practitioner/training
- AI and cybersecurity practitioners play differing roles in risk management and use differing terminology

Cyber AI Profile Focus

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Organizations vary on whether and how they are using AI in these three areas. Some organizations may not yet be using AI. Regardless of where they are on their AI journey, organizations need risk management approaches that support the realities of advancements in AI use.

Securing AI ***in*** the enterprise (***Secure***)

Challenges in Protecting Enterprise AI

- Data Governance, Security, and Privacy
- Subject to Adversarial Attacks
- Unauthorized Access and Use
- Identifying, Detecting, & Responding to AI Component Vulnerabilities and Adverse Events
- Supply Chain Security
- Model Drift & Unexpected or Inaccurate Results

Using AI to protect the enterprise (***Defend***)

Opportunities for AI in Cyber Defense

- Advanced Threat Detection
- Advanced Threat Analysis
- Automated Incident Response
- Proactive Risk Management
- Security Governance & Policy

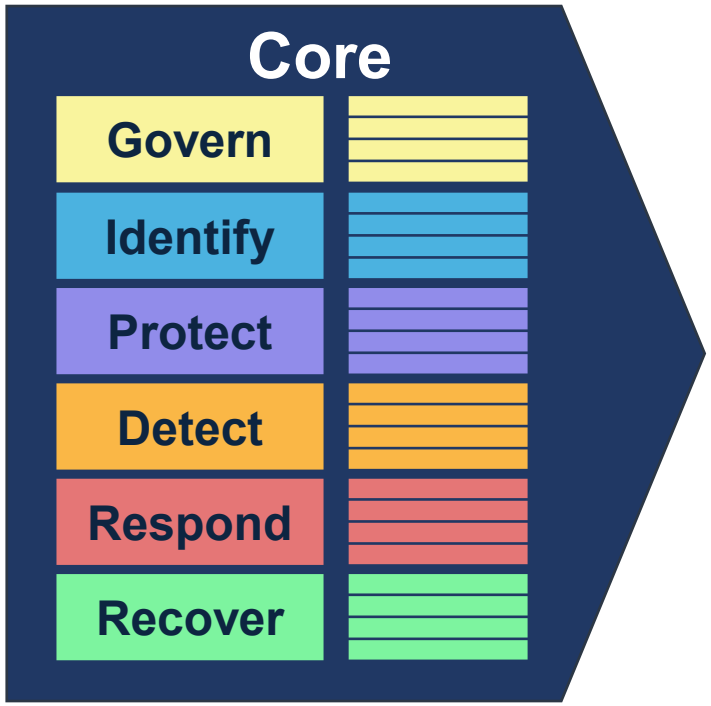
Protecting the enterprise ***from*** AI (***Thwart***)

New or Augmented Threats

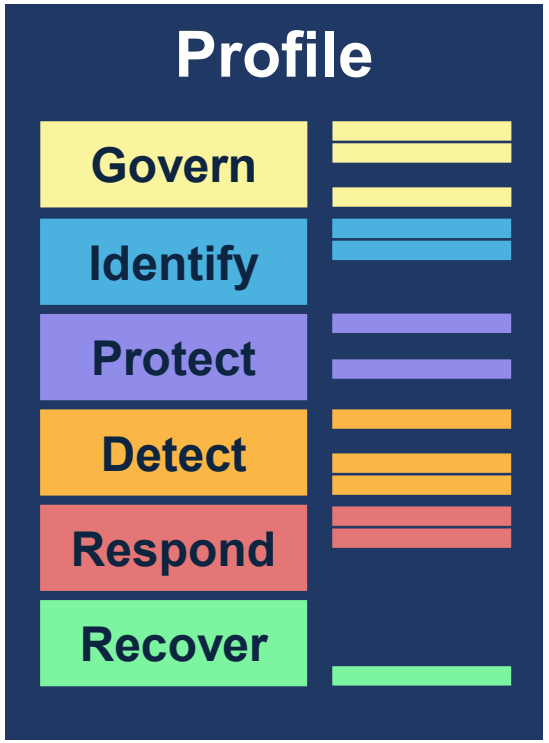
- Automated & Adaptive Malware
- Targeted Phishing & Social Engineering
- AI-Driven Cyber Espionnage
- Evasion Techniques
- Supply Chain Attacks
- AI-Powered Zero-Day Exploits
- AI-Powered Attack Automation

For more information:
https://www.nccoe.nist.gov/projects/cyber-ai-profile

About CSF Profiles



The Core presents all cybersecurity outcomes from the CSF



A Profile can present prioritized, shared expectations regarding community-specific or technology specific considerations and needs

What could be in a Cyber AI Profile?

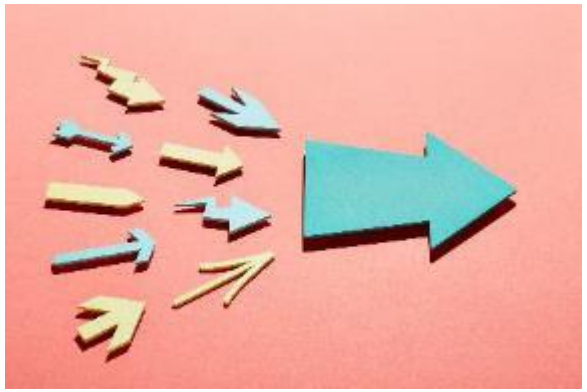
NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

The outcomes described in the NIST Cybersecurity Framework (CSF) 2.0 provide a practical way to help organizations understand, examine, and address the cybersecurity risks introduced by the adoption of AI.

Common Priorities

AI-specific Cybersecurity Implications

Illustrative Examples and Informative References



Mappings to Other NIST Frameworks

Align Risk Considerations to NIST CSF 2.0

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

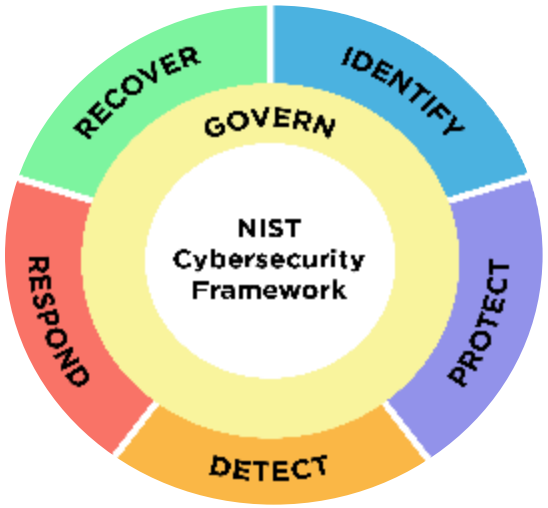
Step 1:
Examine available publications and data collected from meetings

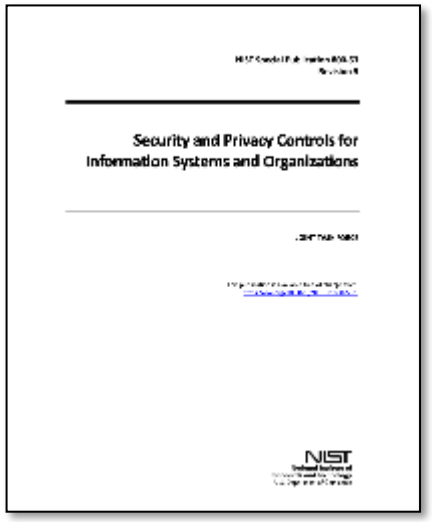
Step 2:
Assess how the identified threats and mitigations are addressed by CSF 2.0

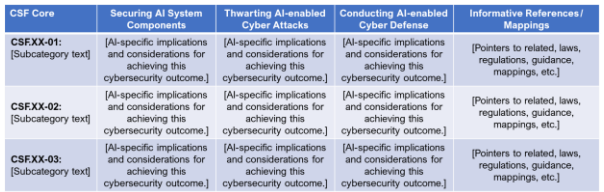
Step 3:
Document unique AI considerations for risk management

Sources of Example Inputs

Concept Documents	Mapped Documents
• Cloud Security Alliance (CSA) • Center for Security and Emerging Technology (CSET) • Institute for Security + Technology (IST) • R Street	• Databricks • European Union Agency for Cybersecurity (enisa) • Google • MITRE ATLAS™ • OWASP




Security and Privacy Controls for Information Systems and Organizations



CSF Core	Securing AI System Components	Thwarting AI-enabled Cyber Attacks	Conducting AI-enabled Cyber Defense	Informative References / Mappings
CSF.XX-01: [Subcategory text]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[Pointers to related, laws, regulations, guidance, mappings, etc.]
CSF.XX-02: [Subcategory text]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[Pointers to related, laws, regulations, guidance, mappings, etc.]
CSF.XX-03: [Subcategory text]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[AI-specific implications and considerations for achieving this cybersecurity outcome.]	[Pointers to related, laws, regulations, guidance, mappings, etc.]

Overall Outcomes of COI Engagement

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Concept Paper

- Confirmed the need for guidelines
- Agreed the NIST CSF was the right starting point
- Identified key challenges

Workshop #1

- Validated direction of the Profile
- Identified critical topics
- Proposed priorities and resources

Summer 2025 COI Working Sessions

- Refined cybersecurity priorities for each Focus Area

Workshop #2

- Continued refining Profile direction and content
- Shared insights regarding implementation

Preliminary Draft Comments

- Informed continued build out of Profile
- Raised areas for further engagement with the COI

Workshop #1 Blog Post

Workshop #2 Blog Post

Workshop #1 Themes

Reflections from Workshop #1

COI Working Session Takeaways

Overall Themes:

- AI risk management requires a multi-disciplinary approach
- AI magnifies some long-standing challenges (e.g., understanding the data landscape within organizations)
- AI readiness (generally) is an important consideration for organizations
- Lack of common taxonomy
- Strong need for risk assessment guidance

COI Working Sessions

Secure

- Priority on governance and risk management (especially risk assessments)
- Importance of cross-functional AI and cybersecurity team
- Tension between accelerating adoption of AI and unclear ownership for managing AI risk
- Need more transparency in supply chain system
- Difficult to protect assets in a dynamic environment

Defend

- Many opportunities to capitalize on AI capabilities
- Need to understand how 3rd party tools were trained
- Challenge identifying when AI is embedded and used
- Need for visibility through monitoring and logging
- Incorporate security controls into AI prompts
- Primary areas of interest: advanced threat detection and analysis, proactive risk management, and security governance & policy

Thwart

- Challenges responding based on speed and scale of attacks
- Introducing new and novel attacks
- Interest in sharing AI cyber threat intelligence
- Identities are critical (human to machine, and machine to machine)
- Continuous awareness and training are essential to thwarting future/similar attacks,
- Technical solutions can be defeated by other technologies, especially AI technologies
- Understanding whether an attack is AI-enabled informs how it is addressed

Workshop #2 Themes

Trustworthiness for AI Decisions

Testing and Evaluation

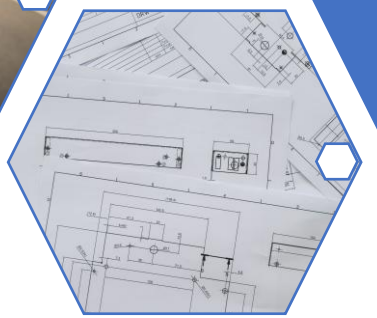
Format and Readability

Consistent AI Taxonomy

Transparency, Integrity, and Accountability

Human-in-the-Loop

Profile Longevity



Use Cases

Reflections from Workshop #2

Preliminary Draft Comments

Preliminary Draft

- >1,400 comments received from 144 submissions:
  - Over half of these offer technical comments on Subcategory considerations
  - The remainder address various other aspects of the preliminary draft

- Examples of themes emerging from the review:
  - Clarifying Focus Area descriptions and relationships
  - Understanding and adjusting proposed priorities
  - Explaining what we meant by “Standard cybersecurity practices apply.”
  - Drawing out unique considerations for Agentic AI
  - Addressing AI Bill of Materials (AIBOM) and Supply Chain considerations
  - Providing the material in a more flexible format (e.g., ability to sort the tables)

Cyber AI Profile Roadmap

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Initial Research and Scope Shaping ☑

NIST collaboration throughout

Literature review continues as needed

Blog Post / Announcement ☑ F

~1000 comments

Cyber AI Profile Concept Paper ☑ F

~400 participants

Public Workshop #1 ☑ F

~350 attendees per session

COI Working Sessions Summer 2025 ☑ F

~1400 comments

Cyber AI Profile Preliminary Draft ☑ F

~2000 participants

Public Workshop #2 ☑ F

~9,400 COI members

COI Working Sessions Spring 2026 ☑ F

Cyber AI Profile Initial Public Draft (IPD) ☐ F

Public Workshop #3 ☐ F

2PD?

Cyber AI Profile FINAL ☐

Organizational Use + Additional Guidance (TBD)

F Opportunities for COI/public stakeholder feedback (NOTE: Internal NIST collaboration occurs throughout)

NIST NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

https://www.nccoe.nist.gov/projects/cyber-ai-profile

CyberAIProfile@nist.gov

nccoe.nist.gov

@NISTcyber