NIST Internal Report
NIST IR 8578

Workshop Summary Report for "Cyber AI Profile" Hybrid Workshop #1

Katerina N. Megas
Bronwyn Patrick
Julie Nethery Snyder

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8578

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Internal Report
NIST IR 8578

Workshop Summary Report for "Cyber AI Profile" Hybrid Workshop #1

Katerina N. Megas
Applied Cybersecurity Division
Information Technology Laboratory

Bronwyn Patrick
Julie Nethery Snyder
National Cybersecurity Center of Excellence
MITRE

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8578

August 2026

U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Dr. Arvind Raman, Under Secretary of Commerce for Standards and Technology and NIST Director

Publication History
Approved by the NIST Editorial Review Board on 2026-07-14

How to Cite this NIST Technical Series Publication
Megas K, Patrick B, Snyder J (2026) Workshop Summary Report for "Cyber AI Profile" Hybrid Workshop. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Internal Report (NIST IR) NIST IR 8578. https://doi.org/10.6028/NIST.IR.8578

Author ORCID iDs
Katerina Megas: 0000-0002-2815-5448
Bronwyn Patrick: 0009-0001-7885-4773
Julie Snyder: 0009-0004-6352-2831

Contact Information
cyberaiprofile@nist.gov

National Institute of Standards and Technology
Attn: Applied Cybersecurity Division, Information Technology Laboratory
100 Bureau Drive (Mail Stop 2000) Gaithersburg, MD 20899-2000

Additional Information
Additional information about this publication is available at https://csrc.nist.gov/pubs/ir/8578/final, including related content, potential updates, and document history.

Abstract

This report summarizes the discussion at the first "Cyber AI Profile Workshop" in April 2025 during which NIST sought feedback toward the development of a NIST Cybersecurity Framework Community Profile for AI. The Cyber AI Profile will provide guidelines for managing cybersecurity risks associated with AI systems and security capabilities. Discussions at the workshop focused on governance and operational challenges organizations face when adopting AI. It built on comments received in response to the February 2025 Cybersecurity and AI Workshop Concept Paper.

Keywords

Artificial Intelligence; Community Profile; Cyber AI Profile; Cybersecurity Framework

Reports on Computer Systems Technology

The Information Technology Laboratory (ITL) at the National Institute of Standards and Technology (NIST) promotes the U.S. economy and public welfare by providing technical leadership for the Nation's measurement and standards infrastructure. ITL develops tests, test methods, reference data, proof-of-concept implementations, and technical analyses to advance the development and productive use of information technology. ITL's responsibilities include developing management, administrative, technical, and physical standards and guidelines for the cost-effective security and privacy of information other than national security-related information in federal information systems.

Table of Contents

List of Tables

Acknowledgments

The authors extend their gratitude to all those who contributed insights that shaped discussions at the workshop. Thank you to the many stakeholders across sectors, the participants, speakers, and moderators of the Cyber AI Profile Workshop, and our colleagues at NIST and The MITRE Corporation, who provided the expertise necessary to conduct this workshop. Additionally, the thoughtful feedback stakeholders shared leading up to the event, including comments on the Cybersecurity and AI Workshop Concept Paper, was instrumental in shaping discussions at the workshop and this document summarizing the workshop.

1. Introduction

On April 3, 2025, the National Institute of Standards and Technology (NIST) National Cybersecurity Center of Excellence hosted a public workshop titled *Cyber AI Profile Workshop*. The event brought together stakeholders from government, industry, and academia to discuss questions in the Cybersecurity and AI Workshop Concept Paper [1] published on February 14, 2025, as well as comments received on the concept paper. This document provides a summary of the workshop's discussions, including opening remarks, fireside chat, panels, and breakout sessions. These activities aimed to define Focus Areas, identify anticipated uses, and outline useful elements for the proposed Cyber Artificial Intelligence (AI) Profile. By incorporating insights from participants across various organizations, the workshop established a collaborative foundation for addressing AI-specific cybersecurity challenges while enabling organizations to harness AI's transformative potential.

1.1. About the NIST Cyber AI Profile

Recent advances in AI technologies bring great opportunities to organizations, but also new and modified risks and impacts that need to be managed in the domain of cybersecurity. NIST is evaluating how to use existing frameworks, such as the NIST Cybersecurity Framework 2.0 (CSF 2.0) [2], to assist organizations as they face these new or expanded risks. Feedback from the cybersecurity community strongly suggests that developing guidelines based on CSF 2.0 to address AI-related cybersecurity risks would be valuable.

Organizations vary in whether and how they use AI in their operations. Some organizations may not yet be using AI. Some may be using cybersecurity solutions enabled with machine learning (ML) technology for pattern detection and planning, but have not yet adopted newer, transformational AI capabilities, such as Generative AI. Regardless of where organizations are on their AI journey, they need risk management approaches that account for the realities of advancements in the use of AI to position them to defend against AI-enabled cyberattacks from adversaries and to leverage AI-enabled cyber defense capabilities. To address these emerging risks and opportunities, particularly in a rapidly evolving field, NIST is developing the Cyber AI Profile to identify how organizations can leverage and adapt the CSF 2.0 [2] to manage AI-related risks effectively [1]. The Cyber AI Profile will seek to help organizations integrate AI-specific considerations into existing cybersecurity programs by:

- Establishing a shared understanding of AI-related cybersecurity priorities and considerations for any organization;
- Fostering collaboration and communication across the AI and cybersecurity communities; and
- Providing common AI cybersecurity target outcomes that organizations can use to support strategic planning efforts and cybersecurity assessments, based on the CSF Core.

1.2. Background for the Cyber AI Profile Workshop

The workshop's goal was to gather community feedback on the Cybersecurity and AI Workshop Concept Paper [1] to inform the direction and contents of the Cyber AI Profile. This concept paper outlined NIST's initial observations and proposed approaches for addressing the intersection of AI and cybersecurity. Recognizing the transformative opportunities and risks introduced by AI, NIST sought input on whether existing frameworks, such as the CSF 2.0 [2] and the Artificial Intelligence Risk Management Framework (AI RMF) [3], could be adapted to address AI-specific challenges effectively.

This workshop was the first workshop, representing an early step in engaging stakeholders to define the scope of the Cyber AI Profile. The workshop explored the scope, focus areas, and practical applications of the Cyber AI Profile, with a focus on securing AI system components that leverage AI for cyber defense and thwarting AI-enabled cyber attacks. By engaging stakeholders across sectors, NIST aimed to ensure the development of meaningful guidelines that support organizations in managing AI-related cybersecurity risks while capitalizing on AI's capabilities. Public feedback on the Cybersecurity and AI Workshop Concept Paper [1] informed the agenda and planned discussions for the Cyber AI Profile Workshop.

1.3. About the Cyber AI Profile Workshop

The agenda for the one-day free, public, hybrid workshop is provided in Table 1. The workshop consisted of a presentation, fireside chat, and two panel discussions that were open to all in-person and livestream audiences, followed by three rounds of in-person breakout sessions related to the cybersecurity challenges AI introduces to organizations and how the Cyber AI Profile could be used to address those challenges.

Table 1. Cyber AI Profile Workshop Agenda

Time	Activity and Presenters
9:00 – 9:10	**Welcome and Opening Remarks: James St. Pierre,** Deputy Director of the NIST Information Technology Laboratory (ITL)
9:10 – 9:20	**NCCoE Welcome: Cherilyn Pascoe,** Director of the National Cybersecurity Center of Excellence
9:20 – 9:50	**Setting the Stage for the Cyber AI Profile: Katerina Megas,** Program Manager for NIST's Cybersecurity, Privacy, and AI Programs
9:50 – 10:50	**Fireside Chat: Potential Interplays Between the CSF and AI RMF for Cybersecurity** • **Moderator:** Dan Caprio, DLA • Stephen Quinn, NIST • Martin Stanley, NIST
10:20 – 11:05	**Panel: Protection of AI Systems** • **Moderator:** Vicky Pillitteri, NIST • Johann Dettweiler, stackArmor • Arun Pamulapati, Databricks

Time	Activity and Presenters
	• Charley Snyder, Google • Sam Wasko, ProtectAI
11:05 – 11:50	**Panel: Defensive and Adversary Use of AI** • **Moderator:** Martin Stanley, NIST • Drew Bagley, CrowdStrike • Dan Kent, Cloudflare • Michelle Sahar, OpenPolicy • Robert Sandler, Trend Micro
1:30 – 4:35	**Breakout Sessions** • Facilitator Track A: Hillary Tran, NIST NCCoE/MITRE • Facilitator Track B: Jon Davis, NIST NCCoE/MITRE • Facilitator Track C: John Dombrowski, NIST NCCoE/MITRE
4:40 – 5:00	**Workshop Close-out and Next Steps** • Katerina Megas, NIST • Hillary Tran, NIST NCCoE/MITRE • Jon Davis, NIST NCCoE/MITRE • John Dombrowski, NIST NCCoE/MITRE

The workshop provided a platform for exploring AI cybersecurity opportunities and challenges, as well as the potential role of the Cyber AI Profile in addressing these issues. Throughout the day, participants engaged in dynamic conversations about governance, framework integration, operationalization challenges, and risk management strategies for AI systems. The workshop highlighted the importance of collaboration between the AI and cybersecurity communities, actionable implementation guidelines, and tailored approaches to securing AI components, thwarting AI-enabled attacks, and leveraging AI for cyber-defense.

Materials from the workshop are available on the event page (https://www.nccoe.nist.gov/get-involved/attend-events/cyber-ai-profile-workshop). Post-workshop materials include a recording of the morning sessions and the presentation slides for the morning sessions and afternoon breakouts.[1]

2. Event Summary

Participants in the workshop explored the cybersecurity opportunities and complex challenges introduced by AI systems, emphasizing the need for actionable solutions and cross-disciplinary collaboration. Discussions focused on integrating AI-specific risks into existing frameworks, such as the NIST CSF 2.0 [2], while addressing governance, operationalization, and risk management strategies. The subsections that follow summarize each presentation and discussion on the agenda.

[1] The afternoon breakout sessions were not recorded.

2.1. Workshop Agenda & Highlights

The workshop began with opening remarks from Jim St. Pierre, Deputy Director of the Information Technology Laboratory (ITL) at NIST, Cherilyn Pascoe, Director of the NCCoE, and Katerina Megas, Program Manager for the NIST Cybersecurity, Privacy, and AI Programs. Each speaker emphasized:

- The importance of collaboration in addressing the intersection of cybersecurity and AI;
- NIST's mission to foster innovation and competitiveness through standards and technology; and
- The critical role of stakeholder input in shaping practical, impactful guidelines.

2.1.1. Welcome and Opening Remarks

Jim St. Pierre, Deputy Director of NIST's ITL, welcomed attendees to the NCCoE and expressed gratitude for their collaboration. He provided an overview of NIST's mission to promote innovation and industrial competitiveness through measurement science, standards, and technology. Highlighting NIST's non-regulatory status, he emphasized NIST's ability to foster close industry collaboration without regulatory constraints. He also introduced ITL's research areas, including cybersecurity, quantum computing, AI, and biometrics, and emphasized the importance of securing AI systems to align with NIST's mission. Mr. St. Pierre concluded by thanking attendees for their input to help shape NIST's outputs.

2.1.2. NCCoE Welcome

Cherilyn Pascoe, Director of the NCCoE, welcomed participants and highlighted the center's collaborative approach to addressing cybersecurity challenges. She emphasized the NCCoE's role in transitioning NIST's world-class research into practical, real-world security solutions. Ms. Pascoe discussed the critical importance of cybersecurity to national and economic security and acknowledged the rapidly evolving landscape of threats and technologies, particularly AI. She outlined the NCCoE's expertise in cybersecurity and AI, including initiatives such as Community Profiles, Privacy-Enhancing Technologies, and the Dioptra testbed. Ms. Pascoe encouraged active participation in the workshop, stressing the importance of community input in developing impactful guidelines.

2.1.3. Setting the Stage for the Cyber AI Profile

Katerina Megas, Program Manager for NIST's Cybersecurity, Privacy, and AI Programs, provided an overview of the program, examples of ongoing and past work at NIST supporting the program's mission, as well as the goals and structure of the workshop. A few examples of current NIST work within the AI space include the AI RMF, the Dioptra testbed, the Adversarial Machine Learning Taxonomy and Terminology of Attacks and Mitigations, the Secure Software Development Framework (SSDF), TrojAI, and the expansion of the NICE Framework (to include Competency Areas aligned to the AI RMF and CSF 2.0). She shared insights from preliminary stakeholder engagements, highlighting three key takeaways: the urgent need for AI-focused

cybersecurity guidelines, the importance of leveraging existing frameworks, and the necessity of a common taxonomy to address cybersecurity challenges in AI.

Ms. Megas introduced the concept of a Cyber AI Profile, which aims to bridge the AI and cybersecurity communities, align requirements, and provide practical guidelines for organizations. She outlined the roadmap for developing the profile, emphasizing collaboration and community input. Ms. Megas also previewed the workshop's breakout sessions, which focused on refining the Profile's Focus Areas, understanding how organizations would leverage the Profile, and what elements the community would like the Profile to contain.

2.2. Fireside Chat: Potential Interplays Between the CSF and AI RMF for Cybersecurity

The fireside chat featured NIST experts Steve Quinn and Martin Stanley, moderated by Dan Caprio, to discuss the interplay between the CSF 2.0 [2] and the AI RMF [3]. The discussion emphasized the need for executive-level governance and oversight to manage AI-related cybersecurity risks effectively. The NIST experts highlighted the importance of integrating AI into enterprise risk management strategies and aligning AI-specific risks with existing cybersecurity frameworks. They also explored the challenges of operationalizing AI risk management and the need for actionable guidelines to help organizations implement AI-enabled cybersecurity controls. Discussions also emphasized the importance of collaboration between AI and cybersecurity communities to address shared challenges and opportunities.

Themes discussed during the fireside chat were grouped by CSF Functions:

Table 2. Themes Discussed During the Fireside Chat by CSF Function

Function	Theme
GOVERN (GV)	• Provide executive level oversight of AI risk to address today's needs and prepare for additional AI capabilities in the future. • Account for upside/positive risks, not just negative risks. • Leverage frameworks to support incorporating AI risk management into current enterprise risk management processes. • Use standardized terminology/taxonomy between AI and cyber communities and practitioners to enable these communities to effectively collaborate. • Apply the forthcoming Cyber AI Profile, which will synthesize elements from NIST frameworks (e.g., CSF, AI RMF) and other resources, to inform a comprehensive cyber-AI strategy.
IDENTIFY (ID)	• Identify risks associated with the adoption of new AI technology.
PROTECT (PR)	• Develop implementation guidance for cybersecurity controls for AI systems.
DETECT (DE)	• Integrate detection of AI threats into current cyber detection capabilities.
RESPOND (RS)	Respond was not directly addressed during the fireside chat.
RECOVER (RC)	Recover was not directly addressed during the fireside chat.

Other considerations mentioned include:

- There is a need to translate existing resources (e.g., frameworks, controls) into actionable implementation guidance for AI rather than creating something wholly new.
- Frameworks can inform executive-level decision-making by providing clarity on acceptable risk levels and strategies for mitigating AI-related threats.
- Frameworks can put risk in perspective by providing clarity on acceptable risk levels and strategies for mitigating AI-related threats.
- Reference architectures and taxonomies are needed.

2.3. Summary of Panel Sessions

2.3.1. Panel 1 – Protection of AI Systems

The first panel of industry experts, moderated by Vicky Pillitteri, focused on identifying the scope, usage, gaps, and next steps for protecting AI systems from a cybersecurity perspective. Panelists highlighted that AI systems consist of multiple components, including data, models, applications, and infrastructure, each requiring tailored security measures. They noted that traditional cybersecurity concepts apply to AI systems, but there are differences in implementation that need to be addressed. Panelists also discussed the importance of understanding and applying current publications and references rather than reinventing the wheel or creating additional guidance, noting that duplicative guidance can be a burden for organizations when addressing threat modeling, supply chain security, and shared responsibility among stakeholders. Panelists highlighted the need for actionable implementation guidelines to address AI-specific risks, such as prompt injection and data poisoning, while leveraging existing cybersecurity principles, as well as the potential need for use case or industry-specific Profiles based on the Cyber AI Profile (after it is published). The discussion also stressed the importance of education and collaboration across teams to bridge knowledge gaps between AI researchers and cybersecurity professionals.

Themes discussed during the panel were grouped by CSF Functions:

Table 3. Themes Discussed During Panel 1 by CSF Function

Function	Theme
GOVERN (GV)	• Understand and manage risks associated with AI supply chain including the data.
IDENTIFY (ID)	• Identify the multiple components of an AI system (model, data, applications) that need to be addressed as part of a comprehensive cybersecurity strategy for AI. • Conduct threat modeling for the entire system and its activities • Understand the provenance of assets. • Understand the risk associated with third party models.

Function	Theme
PROTECT (PR)	• Develop implementation guidance specific to AI (e.g., prompt injection and data poisoning). • Provide training throughout the organization to understand proper use of AI and manage AI-related risks. • Protect data (both protection of sensitive data and privacy are concerns) • Scan assets for vulnerabilities.
DETECT (DE)	Detect was not directly addressed during this panel.
RESPOND (RS)	Respond was not directly addressed during this panel.
RECOVER (RC)	Recover was not directly addressed during this panel.

2.3.2. Panel 2 – Defensive and Adversarial Use of AI

The second panel of industry experts, moderated by Martin Stanley, explored the use of AI in cybersecurity defense and the adversarial use of AI to augment cyberattacks. Panelists discussed how AI enhances cybersecurity tools by improving threat detection, incident response, and behavioral analysis. They highlighted the role of AI in enabling predictive path analysis and reducing alert fatigue for defenders. On the adversarial side, panelists noted that AI lowers the barrier for attackers, enabling faster and more scalable attacks, such as AI-crafted phishing campaigns and automated vulnerability exploitation. The discussion emphasized the need for defenders to leverage AI to match adversaries' speed and scale while addressing transparency, trust, and accountability in AI-enabled systems. Panelists called for red teaming, continuous risk evaluation, governance frameworks, and robust identity management to mitigate risks associated with adversarial machine learning and protecting AI systems after adoption. Panelists also acknowledged that there may be a need for additional regulation in this space to address any identified gaps.

Themes discussed during the panel were grouped by CSF Functions:

Table 4. Themes Discussed During Panel 2 by CSF Function

Function	Theme
GOVERN (GV)	• Expand governance practices to understand the role of AI in the organization and to address its intersections with other areas (e.g., privacy). • Understand the opportunities and risks associated with the use of AI to defend systems. • Develop a methodology to gain insight into decision making by AI systems.
IDENTIFY (ID)	• Know the organization's points of vulnerability throughout the network and the data lifecycle. Inventory AI assets (e.g., hardware, models, application programing interfaces (APIs), keys, agents).

Function	Theme
	• Take a layered approach to thwarting against attacks, factoring in networks, containers, servers, platforms, data, and infrastructure. • Understand the risks associated with attackers leveraging AI using existing attack vectors. • Recognize both common and unique aspects of AI risks. • Understand risks to supply chain of AI tools. • Use simulation capabilities to better prepare against adversarial tools. • Scale identity management practices to address how quickly and easily adversaries can move through a compromised network. • Build AI considerations into red-teaming practices. • Incorporate feedback loops to continue improving AI use in cybersecurity activities.
PROTECT (PR)	• Capitalize on the potential of AI capabilities to strengthen defenses. • Recognize the risks of adversarial AI to compromise identity management. • Understand the use cases for and protections necessary to effectively use agentic AI.
DETECT (DE)	• Capitalize on new monitoring and analysis capabilities (e.g., user and machine behavioral analysis, real-time monitoring and response). • Leverage AI during event analysis. For example, it is already frequently used to increase the efficiency of activities like anomaly detection and it can enhance an organization's ability to conduct more holistic analyses using inputs from multiple sources (different types of devices and telemetries). • Identify AI capabilities that can support organizations in moving from a reactive defensive mode to proactive (e.g., predictive analysis).
RESPOND (RS)	• Enhance response and recovery efforts, for example using AI tools to recommend steps to take based on an event. • Understand the risk associated with using AI for response.
RECOVER (RC)	Recover was not directly addressed during this panel.

2.4. Summary of Breakout Sessions

2.4.1. Track A – Focus Area Descriptions

This breakout session focused on refining the descriptions of the three primary Focus Areas for the Cyber AI Profile: securing AI system components, thwarting AI-enabled cyberattacks, and using AI for cyber-defense activities. Characteristics that arose for each Cyber AI Profile Focus Area are listed below.

Securing AI System Components:

Participants emphasized the expanded attack surface introduced by AI systems, including risks tied to data, models, infrastructure, and applications. Discussions highlighted the dynamic nature of AI systems, their interaction with external data sources, and supply chain dependencies, as well as governance challenges related to contracts, regulations, and organizational culture shifts.

Conducting AI-Enabled Cyber Defense:

The session highlighted efficiency gains in tasks like anomaly detection and incident response while addressing risks such as false positives/negatives. Participants also stressed the importance of benchmarking AI performance and integrating multidisciplinary perspectives to ensure effective implementation.

Thwarting AI-Enabled Cyber Attacks:

Participants explored the scale, speed, and automation of such AI-enabled threats, advocating for proactive defense measures, incident response plans tailored to AI-specific risks, and the adoption of zero-trust architecture principles.

CSF-related topics that arose throughout the Focus Area discussions were grouped by CSF Functions and included:

Table 5. Themes Discussed During Track A Focus Group Breakouts by CSF Function

Function	Theme
GOVERN (GV)	• Explainability and confidence levels in AI interactions are critical for trustworthiness and regulatory compliance. • Transparency in the AI supply chain is essential. • Need for metrics to assess AI defense capabilities. • Need to understand the relationship between AI risks and privacy risks. • Use contracts to understand and govern AI-related cybersecurity requirements.
IDENTIFY (ID)	• Identify risks associated with the use of AI systems, including third-party and supply chain risks (e.g., AI integrated into purchased applications). • Consider security throughout the data life cycle. • Need to understand data quality and use (generally and for cybersecurity activities) and the potential impacts to privacy of individuals (e.g., ensuring data is not leaked from a model).
PROTECT (PR)	• Strong access controls (e.g., access control lists, role-based access controls) to accomplish zero trust architectures (ZTA) will be necessary to manage access to data. • The workforce will need role-based training and awareness to understand cybersecurity implications related to AI and to ensure proper use.

Function	Theme
	• Implement cybersecurity controls for AI that scale. • Protect against prompt injections and data poisoning. • Protect against supply chain attacks.
DETECT (DE)	• Develop novel methods needed to detect prompt injection and data poisoning. • Mitigate false positives generated by AI hallucinations.
RESPOND (RS)	• Address AI-specific considerations in incident response plans.
RECOVER (RC)	Recover was not directly addressed during Track A.

2.4.2. Track B – Anticipated Profile Uses and Elements

This session explored the anticipated uses of the Cyber AI Profile and the essential elements it should include.

Ways participants envision using the Cyber AI Profile include:

- **Supporting Effective Communication:** Participants view the Cyber AI Profile as a tool to communicate an organization's risk posture and provide metrics that customers, partners, and regulators can rely on.
- **Aligning the Landscape:** Participants plan to use the Cyber AI Profile to harmonize across the growing body of standards of guidance.

Participants also expressed the following additional desired characteristics of the Cyber AI Profile:

- **Scalability:** Participants noted that the Cyber AI Profile should be scalable to meet the needs of any size organization.
- **Common taxonomy:** Participants view the Cyber AI Profile as a way to create a governance taxonomy between the AI and cybersecurity communities, which can be used to describe the relationship between cybersecurity and AI risk. Participants noted the connections between existing frameworks, including the NIST Cybersecurity Framework (CSF), NIST AI Risk Management Framework (AI RMF), NIST Privacy Framework, NIST Risk Management Framework (NIST RMF), and Security and Privacy Controls for Information Systems and Organizations.
- **Operational integration:** Participants emphasized the need for practical implementation guidance that organizations could operationalize within existing security processes.
- **Harmonization:** Identify existing guidance and understand how these emerging guidance documents can be used to develop a standard approach to cybersecurity and AI risk management. Suggestions included mappings to international standards, such as the European Union's General Data Protection Regulation (GDPR) and the anticipated ISO/IEC 27090, and leveraging tools like MITRE ATLAS™ for threat modeling.

- **Interactive and Customizable:** Participants would like the Cyber AI Profile to be produced in a format that enables users to interact with it during use, including features such as weighting Categories and Subcategories to facilitate prioritization and to provide more specific information to supplement the Core.
- **Metrics:** Participants expressed interest in modular and repeatable metrics.
- **Future-proof:** Participants emphasized the importance of future-proofing the Profile to accommodate emerging technologies and the evolving regulatory and standards landscape.

While the purpose of this session was to understand the use of the Cyber AI Profile and the types of content the Profile should contain for usability, the discussions also yielded several insights into specific considerations related to the CSF Core:

Table 6. Themes Discussed During Track B Focus Group Breakouts by CSF Function

Function	Theme
GOVERN (GV)	• Describe oversight-level considerations, particularly for boards of organizations. • Address responsibility for managing risks based on various types of AI systems the organization is adopting/implementing (e.g., Large Language Models (LLMs,) generative AI.
IDENTIFY (ID)	Identify was not directly addressed during Track B.
PROTECT (PR)	• Need to focus on data integrity in addition to data security • Emphasize AI's significant dependency on data governance, potentially as its own Profile.[2]
DETECT (DE)	Detect was not directly addressed during Track B.
RESPOND (RS)	Respond was not directly addressed during Track B.
RECOVER (RC)	Recover was not directly addressed during Track B.

2.4.3. Track C – Priorities in the CSF Core

This session used the Functions and Categories of the CSF Core to begin identifying the top threats and unique considerations for the three Cyber AI Profile Focus Areas outlined in the concept paper. Participants emphasized the need to integrate AI-specific risks into existing frameworks and to provide actionable guidelines for organizations; they indicated they like the idea of treating the three Focus Areas separately to address their distinct challenges while recognizing areas of overlap, such as governance and monitoring. The table below provides an overview of themes that arose for each Focus Area by CSF Function and reflects cross-cutting themes across all three Focus Areas.[3]

[2] See this related NIST Data Governance and Management Profile effort.
[3] The table only reflects topics that were discussed during the short time allocated for each breakout session. This table is not intended to provide a comprehensive list of considerations. This topic was further explored in a series of virtual Cyber AI Profile Community of Interest working sessions.

Table 7.: Themes Discussed for Each Focus Area During Track C Focus Group Breakouts by CSF Function

Function	Securing AI System Components	Conducting AI-enabled Cyber Defense	Thwarting AI-enabled Cyber Attacks
GOVERN (GV)	• Certification systems for AI components to ensure integrity and reduce risks associated with open-source models and third-party dependencies.	• Need for privacy protection and addressing bias concerns in AI models.	• Emphasis on supply chain security.
	• Emphasis on data governance, privacy, and supply chain security, use of autonomous AI systems, and incident accountability. • Cross-team collaboration to bridge the gap between AI and cybersecurity experts. • Multidisciplinary approach managing AI cybersecurity opportunities and risks, involving relevant parties like legal and procurement/acquisitions. • Opportunities to integrate AI-specific risks into existing frameworks, including feedback loops to improve governance and monitoring.		
IDENTIFY (ID)	• AI systems introduce new system components, which expand the attack surface (e.g., AI agents, public-facing systems, and supply chain dependencies). • Examples of threats identified included: Data poisoning, and supply chain vulnerabilities.	• Lack of standardized benchmarks for measuring AI's effectiveness in cybersecurity. • Emphasis on red teaming/adversarial testing.	• How to differentiate between AI-enabled attacks and non-AI-enabled attacks. • Examples of threats identified included: Adversarial use of AI tools makes existing violations faster and more novel, amplifying risks rather than introducing entirely new ones (e.g., social engineering becomes faster and more realistic and can be done on a larger scale).
	• Tracking data provenance, bias detection in models, and supply chain risks.		
PROTECT (PR)	• Need for robust security measures, data governance,	Specific theme not identified.	Specific theme not identified.

Function	Securing AI System Components	Conducting AI-enabled Cyber Defense	Thwarting AI-enabled Cyber Attacks
	and privacy considerations. • Importance of training and awareness across the organization to address AI-related risks.		
	• Importance of education to bridge the gap between AI and cybersecurity experts. • Cryptographic signing, boundary protection, and prompt injection defenses. • The importance of data integrity, accuracy, and provenance was emphasized, particularly for AI systems reliant on external or third-party data sources.		
DETECT (DE)	Specific theme not identified.	• Examples of opportunities to use AI in cybersecurity defense included: AI enhances efficiency in cybersecurity by analyzing large volumes of logs, detecting threats faster, and saving time and resources.	• Need for controls to catch AI-enabled attacks, such as prompt injection and data poisoning.
	• Automated anomaly detection and continuous monitoring of AI systems. • AI is likely to play a larger role in Detect as the availability of tools that support outcomes in this Function increases (e.g., monitoring).		
RESPOND (RS)	Specific theme not identified.	• Understanding the circumstances around automated responses and when there is a need for human intervention. • Examples of opportunities to use AI in cybersecurity defense included: AI-driven automation and	Specific theme not identified.

Function	Securing AI System Components	Conducting AI-enabled Cyber Defense	Thwarting AI-enabled Cyber Attacks
		threat intelligence sharing.	
	• The importance of considering the autonomous nature of AI and its impact on incident response. • Tailored incident response plans for AI-specific threats.		
RECOVER (RC)	Specific themes not identified.		

3. Key Takeaways

The Cyber AI Profile Workshop offered participants an opportunity to share feedback and insights with NIST throughout the day. These takeaways reflect ideas that resonated strongly across discussions and provide valuable feedback for NIST's ongoing efforts for the Cyber AI Profile and other NIST programs. Participants emphasized the importance of integrating AI-specific risks into existing frameworks, adopting proactive defense strategies, and ensuring supply chain security. Transparency, accountability, and continuous monitoring were identified as critical for maintaining trust and system integrity. Participants also called for the Cyber AI Profile to be modular and adaptable, to include sector-specific Profiles, and to cover the importance of multidisciplinary collaboration to ensure the Profile's relevance across diverse sectors and organizational sizes. Attendees requested that the Cyber AI Profile focus on risk management rather than compliance, providing actionable guidelines to help organizations navigate the evolving cybersecurity landscape shaped by AI technologies.

The summarized takeaways outlined below do not represent formal recommendations to NIST but rather serve as a foundation for future community engagement. These takeaways will inform NIST's continued efforts to develop the Cyber AI Profile as a practical resource for organizations navigating AI-related risks and opportunities.

- **Enterprise Risk Management and Governance:** Participants generally agreed that AI risk is organizational risk. As such, it is critical to integrate AI risk management into existing enterprise risk management practices and governance structures.

- **Multidisciplinary Collaboration and Education:** AI use introduces considerations from multiple aspects of organizational operations, including legal, technical, procurement/acquisitions, and governance teams. Collaboration across these areas was highlighted as essential for addressing AI-related cybersecurity risks. Participants called for integrating perspectives through multidisciplinary teams to ensure a comprehensive enterprise view. They also discussed the need for education and cross-team collaboration to bridge the knowledge gaps between AI and cybersecurity experts.

- **Dual-Use Nature of AI:** AI's dual-use nature, both as a tool for enhancing defenses and as a means for adversaries to amplify attacks, presents both opportunities and risks. Cybersecurity considerations need to address these dual uses. AI can enhance

cybersecurity tools, such as anomaly detection and incident response, but it can also enable adversaries to scale and automate attacks, including phishing, data poisoning, and model inversion. Proactive defense measures, such as automated red teaming and zero-trust principles, were identified as essential.

- **Effectiveness in Cyber-Defense:** AI enhances cybersecurity efficiency, enabling faster data analysis and reduced reliance on third-party services. However, participants stressed the importance of measurable benchmarks to assess AI performance and mitigate challenges such as false positives and negatives, ensuring AI-driven cyber-defense capabilities perform as expected.

- **Supply Chain Security:** The workshop underscored the importance of securing the AI supply chain, including the need to understand the origins of AI components (e.g., microservices, containers, libraries) and data, the new vulnerabilities they may introduce, and their potential impacts on cybersecurity. Discussions also covered open-source models, training and use of third-party models, understanding the impacts of model changes over time, and third-party dependencies (as well as the need for trust and provenance).

- **Transparency:** Participants called for greater transparency in AI systems used to support cybersecurity, including clear documentation of data provenance, model behavior, and decision-making processes. Concerns about the distribution of data leading to incorrect results were identified as areas requiring more focus to ensure trustworthiness in cybersecurity implementations.

- **Human-in-the-Loop:** Participants stressed the importance of human-in-the-loop processes and training to ensure effective implementation and oversight of AI systems used for cybersecurity. Discussions highlighted challenges in understanding AI behavior, including the variability of responses and the difficulty in interpreting AI-driven decisions. Bridging knowledge gaps between AI researchers and cybersecurity professionals was identified as critical for enabling the two communities to work together to build and maintain trustworthy AI systems.

- **Framework Integration:** Participants emphasized the need for a shared understanding and common taxonomy between the AI and cybersecurity communities and noted the importance of integrating AI-specific risks into existing frameworks, such as the NIST Cybersecurity Framework (CSF), NIST AI Risk Management Framework (AI RMF), NIST Privacy Framework, NIST Risk Management Framework (NIST RMF), and Security and Privacy Controls for Information Systems and Organizations. Suggestions included mapping to international standards, such as the European Union's General Data Protection Regulation (GDPR) and the anticipated ISO/IEC 27090, and leveraging other resources, such as MITRE ATLAS™, for threat modeling. The need for sector-specific Community Profiles tailored to industries like healthcare, finance, and energy was also

discussed.

- **Implementation Guidance and Informative References:** Multiple discussions stressed the need for guidance on implementing controls for AI systems and informative references to help organizations understand how to manage AI-related risks. *See this related NIST effort: SP 800-53 Control Overlays for Securing AI Systems (COSAiS).*

- **Data Governance**: Data governance was identified as a critical topic, particularly for AI systems reliant on external or third-party data sources. Participants called for a separate Focus Area to address sensitive data and its use in AI systems, including privacy considerations. *See this related NIST effort: Data Governance and Management Profile.*

Additionally, there was significant emphasis on the need to supplement traditional cybersecurity measures with those tailored for AI. Examples of tailored approaches discussed during the workshop include:

- Cryptographic signing for models.
- Certification systems.
- Use of AI Software Bill of Materials (AI SBOM) to enhance transparency and accountability in AI components.
- Securing feedback loops, where model outputs are reused for further training or decision making, to prevent amplification of errors or adversarial manipulations.
- Robust governance strategies to address risks, such as data poisoning and model manipulation.
- Adaptive identity management and access control strategies and policies to address the growing and complex role of non-human identities, such as AI agents and applications, in organizational environments.
- Incident response strategies and plans that distinguish attacks against AI models from traditional cybersecurity incidents and facilitate appropriate responses to attacks on AI systems (e.g., reverting to last known-good versions of models after model poisoning).
- Data security throughout its lifecycle in AI systems (e.g., model inputs and outputs)

These takeaways provide a foundation for future community engagement.

3.1. Adjacent Considerations

Throughout the event, participants highlighted several important themes that extend beyond the immediate scope of the Cyber AI Profile. While these topics will not be directly addressed in the current iteration of the Profile, they offer valuable insights that may inform future initiatives for the AI Cybersecurity and Privacy Program and other NIST programs. By acknowledging these ideas, NIST reaffirms its commitment to adapting to the evolving needs of the community and exploring opportunities for continued engagement and development.

- **Implementation Based on AI Maturity:** Participants highlighted the need for the Cyber AI Profile to support diverse use cases and organizational contexts. This includes tailoring implementation guidance based on the maturity of AI adoption, such as developing models from scratch, fine-tuning existing models, or using third-party tools.

By addressing these varying levels of maturity, the profile can help organizations apply security measures and governance strategies that align with their specific AI capabilities and risks.

- **Tailored Support for Small and Medium Businesses (SMBs):** Participants emphasized the need for the Cyber AI Profile to provide SMBs with tailored support, including simple acquisition questions and practical tools to address their unique challenges.
- **Sector-Specific Profiles:** Discussions highlighted the desire for NIST to develop sector-specific Cyber AI Profiles and tailored incident response plans to address diverse use cases and ensure relevance across industries.
- **Scalable and Adaptable Tools:** The need for scalable and adaptable tools to manage AI-related cybersecurity risks was emphasized. Participants suggested interactive and customizable design features for the Cyber AI Profile, such as OLIR-based mappings and LLM integration, to make the profile more user-friendly and applicable across diverse industries.
- **SP 800-53 rev 5 Crosswalk:** Participants recommended leveraging NIST SP 800-53 controls to provide detailed, implementable security measures tailored to AI-specific risks.
- **AI Supply Chain Documentation:** Participants advocated for the inclusion of an AI Software Bill of Materials (SBOM) to address supply chain risks associated with AI systems. This initiative would enhance transparency and accountability by identifying the provenance, authenticity, and security of AI components, including open-source models and third-party dependencies. Cryptographic signing and certification systems were suggested as additional measures to ensure the integrity of AI systems across the supply chain.
- **Workforce Development:** Participants emphasized the need for workforce training and development to address AI-related cybersecurity risks and to prepare cybersecurity teams for the effects of adopting AI-enabled tools.

4. Next Steps

The workshop provided NIST with valuable feedback that validated the use of the three focus areas (Secure, Defend, and Thwart) and their definitions, and informed the Preliminary Draft of the Cyber AI Profile. The summer 2025 Community of Interest (COI) Working Sessions[4] followed up on several lines of discussion surfaced at this event. This stakeholder input is essential to the open, transparent process for developing the Cyber AI Profile. Each step of the process toward the Cyber AI Profile will build on the previous actions and information. This workshop was informed by comments received on the Cybersecurity and AI Workshop Concept Paper [1] and the themes gleaned from those responses. The insights from this workshop, as well as comments submitted on the Cyber AI Profile Concept Paper [1], informed the direction of the Cyber AI Profile Preliminary Draft.

[4] Materials from Summer 2025 Community of Interest meetings are available at https://www.nccoe.nist.gov/projects/cyber-ai-profile/2025-virtual-working-sessions-materials

The most current version of the Cyber AI Profile roadmap is maintained here. For more information on updates to the Cyber AI Profile, Cyber AI Profile roadmap phases, or to join the Cyber AI Profile COI, please visit the NCCoE Cyber AI Profile project page, available at https://www.nccoe.nist.gov/projects/cyber-ai-profile.

References

[1] National Institute of Standards and Technology (2025). *Cybersecurity and AI Workshop Concept Paper National Institute of Standards and Technology.* Available at https://www.nccoe.nist.gov/sites/default/files/2025-02/cyber-ai-concept-paper.pdf

[2] National Institute of Standards and Technology (2024). The NIST Cybersecurity Framework (CSF) 2.0 (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) 29. https://doi.org/10.6028/NIST.CSWP.29

[3] National Institute of Standards and Technology (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* Available at https://doi.org/10.6028/NIST.AI.100-1

[4] National Institute of Standards and Technology (2021). NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0 (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) 10. Available at https://doi.org/10.6028/NIST.CSWP.01162020

[5] National Institute of Standards and Technology (2020). *National Online Informative References Program (OLIR).* Available at https://csrc.nist.gov/projects/olir